GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 26, 2003, 19:22:26 ; Search time 246 Seconds
(without alignments)
5922.939 Million cell updates/sec

Title: US-09-677-653A-50
Perfect score: 3374
Sequence: 1 MGDAGVASQRPHNRRGTRNV.....GKIAARVRARRARRRAARAN 647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09677653/runat_20022003_164314_1243/app_query.fasta_1.839
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09677653_@CGN_1_1_157_@runat_20022003_164314_1243 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3374	100.0	2478	19	AAT99117	Helicoverpa armige
	2	3353	99.4	2478	15	AAQ58523	Sequence of Heliot
	3	294	8.7	6536	19	AAV04471	Nudaurelia beta-li
	4	294	8.7	6536	19	AAT99118	Nudaurelia beta vi
	5	139.5	4.1	3695	20	AAX20209	Enterococcus faeca
	6	139.5	4.1	3695	24	ABN98194	E faecalis EF108 g
	7	139.5	4.1	3840	20	AAX20208	Enterococcus faeca
	8	139.5	4.1	3840	24	ABN98193	E faecalis EF108 g
	9	139.5	4.1	19031	20	AAX13104	Enterococcus faeca
	10	130.5	3.9	8496	23	ABL30182	Drosophila melanog
	11	129	3.8	11739	20	AAX13087	Enterococcus faeca
	12	127	3.8	6592	23	ABL19215	Drosophila melanog
	13	126.5	3.7	8277	22	AAS31116	Human diagnostic a
	14	126	3.7	1548	24	ABQ90255	M. capsulatus gene
	15	125.5	3.7	2748	20	AAX81820	Talaromyces emerso
	16	124.5	3.7	7035	23	AAS54978	Staphylococcus aur
	17	124	3.7	29139	23	AAS59569	Propionibacterium
	18	124	3.7	4403765	22	AAI99683	Mycobacterium tube
	19	123.5	3.7	2935	23	AAS88066	DNA encoding novel
c	20	123.5	3.7	265118	22	AAH41227	Pyrococcus abyssi
c	21	122	3.6	47981	22	AAF30757	Micromonospora meg
c	22	121.5	3.6	21500	23	AAS59633	Propionibacterium
	23	121	3.6	1724	24	ABQ68825	Listeria monocytog
	24	121	3.6	2721	6	AAN50114	DNA sequence encod
c	25	121	3.6	4403765	22	AAI99683	Mycobacterium tube
c	26	121	3.6	4411529	22	AAI99682	Mycobacterium tube
	27	120.5	3.6	3410	22	AAD17326	Streptococcus agal
	28	120.5	3.6	4493	20	AAX83610	Human ontherin enc
	29	120.5	3.6	4493	20	AAX77400	Cadherin-like poly
	30	120.5	3.6	38734	20	AAZ32020	Human METH1 relate
	31	120.5	3.6	38734	22	AAC90077	AL021529 cDNA clon
	32	120	3.6	2798	20	AAX20207	Enterococcus faeca
	33	120	3.6	2798	24	ABN98192	E faecalis EF107 g
	34	120	3.6	3078	20	AAX20206	Enterococcus faeca
	35	120	3.6	3078	24	ABN98191	E faecalis EF107 g
	36	119.5	3.5	4278	23	ABL30183	Drosophila melanog
	37	119	3.5	2907	22	AAK52261	Human polynucleoti
	38	119	3.5	5397	23	ABL29757	Drosophila melanog
	39	119	3.5	5982	24	ABQ69372	Listeria innocua D
	40	119	3.5	37856	21	AAA11992	S. cellulosum DNA
	41	119	3.5	1163020	24	ABQ67197	Listeria innocua c
	42	119	3.5	3011208	24	ABQ69245	Listeria innocua D
c	43	118.5	3.5	2034	24	ABQ70603	Listeria monocytog
	44	118	3.5	1590	21	AAF12883	Aspergillus oryzae
	45	117.5	3.5	4776	20	AAZ23707	S. erythraea pFL37

ALIGNMENTS

RESULT 1
AAT99117
ID AAT99117 standard; cDNA; 2478 BP.
XX
AC AAT99117;
XX
DT 08-JUN-1998 (first entry)
XX
DE Helicoverpa armigera RNA2 encoding coat proteins p71 and p17.
XX
KW Vaccine; coat protein; p71; p17; insecticide; Ig-like domain; ds.
XX
OS Helicoverpa armigera.
XX
FH Key Location/Qualifiers
FT CDS 283..756
FT /*tag= a
FT /product= p17 coat protein

```
FT   CDS             366..2309
FT                   /*tag= b
FT                   /product= p71 coat protein
XX
PN   WO9746666-A1.
XX
PD   11-DEC-1997.
XX
PF   02-JUN-1997;   97WO-AU00349.
XX
PR   31-MAY-1996;   96AU-0000234.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Gordon KH,  Hanzlik TN;
XX
DR   WPI; 1998-042175/04.
DR   P-PSDB; AAW34534, AAW34535.
XX
PT   Modified small RNA viruses and virus-like particles - have altered
PT   or substituted Ig-like domains to modify host cell tropism, useful
PT   as insecticides and in medicinal applications
XX
PS   Disclosure; Figure 1; 41pp; English.
XX
CC   The sequence is that of a cDNA encoding RNA2 of the viral genome which
CC   putatively codes for coat proteins p71 and p17. The p71 coat protein
CC   includes an Ig-like domain which can be used in the production of
CC   virus-like particles (VLP). The VLPs can be used in vaccines where the
CC   Ig-like domain has been altered so that the VLP presents a surface
CC   located antigen which is used to elicit an immune response in a host
CC   organism. They can also be used controlling the proliferation of a pest
CC   insect and potentially as medicinal delivery agents for cancer treatment
CC   and gene therapy.
XX
SQ   Sequence 2478 BP; 516 A; 854 C; 607 G; 501 T; 0 other;

Alignment Scores:
Pred. No.:                3.27e-274      Length:          2478
Score:                    3374.00        Matches:         647
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       19             Gaps:            0

US-09-677-653A-50 (1-647) x AAT99117 (1-2478)

Qy       1 MetGlyAspAlaGlyValAlaSerGlnArgProHisAsnArgArgGlyThrArgAsnVal 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 ATGGGAGATGCTGGAGTGGCGTCACAGCGACCTCACAACCGTCGCGGAACCCGTAACGTT 425

Qy      21 ArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArgThrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 CGGGTCAGCGCCAACACCGTCACCGTCAATGGTAGAAGAAACCAACGGCGTCGGACCGGA 485

Qy      41 ArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeu 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     486 AGGCAAGTTTCTCCCCCTGACAATTTCACCGCTGCTGCACAAGACCTCGCGCAAAGCCTT 545

Qy      61 AspAlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrp 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     546 GACGCCAACACCGTCACTTTCCCCGCTAACATCTCTAGCATGCCCGAATTCCGGAATTGG 605

Qy      81 AlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTrpTyrPheLysTyrLeuAsp 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     606 GCCAAGGGAAAGATCGACCTCGACTCCGATTCCATCGGCTGGTACTTCAAGTACCTTGAC 665

Qy     101 ProAlaGlyAlaThrGluSerAlaArgAlaValGlyGluTyrSerLysIleProAspGly 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     666 CCAGCGGGTGCTACAGAGTCTGCGCGCGCCGTCGGCGAGTACTCGAAGATCCCTGACGGC 725

Qy     121 LeuValLysPheSerValAspAlaGluIleArgGluIleTyrAsnGluGluCysProVal 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     726 CTCGTCAAGTTCTCCGTCGACGCAGAGATAAGAGAGATCTATAACGAGGAGTGCCCCGTC 785

Qy     141 ValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPhe 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     786 GTCACTGACGTGTCCGTCCCCCTCGACGGCCGCCAGTGGAGCCTCTCGATTTTCTCCTTT 845

Qy     161 ProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMetSerLeu 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 CCGATGTTCAGAACCGCCTACGTCGCCGTAGCGAACGTCGAGAACAAGGAGATGTCGCTC 905

Qy     181 AspValValAsnAspLeuIleGluTrpLeuAsnAsnLeuAlaAspTrpArgTyrValVal 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     906 GACGTTGTCAACGACCTCATCGAGTGGCTCAACAATCTCGCCGACTGGCGTTATGTCGTT 965

Qy     201 AspSerGluGlnTrpIleAsnPheThrAsnAspThrThrTyrTyrValArgIleArgVal 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 GACTCTGAACAGTGGATTAACTTCACCAATGACACCACGTACTACGTCCGCATCCGCGTT 1025

Qy     221 LeuArgProThrTyrAspValProAspProThrGluGlyLeuValArgThrValSerAsp 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1026 CTACGTCCAACCTACGACGTTCCAGACCCCACAGAGGGCCTTGTTCGCACAGTCTCAGAC 1085

Qy     241 TyrArgLeuThrTyrLysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGln 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 TACCGCCTCACTTATAAGGCGATAACATGTGAAGCCAACATGCCAACACTCGTCGACCAA 1145

Qy     261 GlyPheTrpIleGlyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspVal 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 GGCTTTTGGATCGGCGGCCAGTACGCTCTCACCCCGACTAGCCTACCGCAGTACGACGTC 1205

Qy     281 SerGluAlaTyrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1206 AGCGAGGCCTACGCTCTGCACACTTTGACCTTCGCCAGACCATCCAGCGCCGCTGCACTC 1265

Qy     301 AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrp 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1266 GCGTTTGTGTGGGCAGGTTTGCCACAGGGTGGCACTGCGCCTGCAGGCACTCCAGCCTGG 1325

Qy     321 GluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrProProAla 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1326 GAGCAGGCATCCTCGGGTGGCTACCTCACCTGGCGCCACAACGGTACTACTTTCCCAGCT 1385

Qy     341 GlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspProAsnAsp 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1386 GGCTCCGTTAGCTACGTTCTCCCTGAGGGTTTCGCCCTTGAGCGCTACGACCCGAACGAC 1445

Qy     361 GlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaVal 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1446 GGCTCTTGGACCGACTTCGCTTCCGCAGGAGACACCGTCACTTTCCGGCAGGTCGCCGTC 1505

Qy     381 AspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPheThrVal 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1506 GACGAGGTCGTTGTGACCAACAACCCCGCCGGCGGCGGCAGCGCCCCCACCTTCACCGTG 1565

Qy     401 ArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeuGluThr 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1566 AGAGTGCCCCCTTCAAACGCTTACACCAACACCGTGTTTAGGAACACGCTCTTAGAGACT 1625

Qy     421 ArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrVal 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1626 CGACCCTCCTCTCGTAGGCTCGAACTCCCTATGCCACCTGCTGACTTTGGACAGACGGTC 1685

Qy     441 AlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeuGlyCysTyrLeuVal 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1686 GCCAACAACCCGAAGATCGAGCAGTCGCTTCTTAAAGAAACACTTGGCTGCTATTTGGTC 1745

Qy     461 HisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaVal 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1746 CACTCCAAAATGCGAAACCCCGTTTTCCAGCTCACGCCAGCCAGCTCCTTTGGCGCCGTT 1805

Qy     481 SerPheAsnAsnProGlyTyrGluArgThrArgAspLeuProAspTyrThrGlyIleArg 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1806 TCCTTCAACAATCCGGGTTATGAGCGCACACGCGACCTCCCGGACTACACTGGCATCCGT 1865
```

```
Qy     501 AspSerPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1866 GACTCATTCGACCAGAACATGTCCACCGCTGTGGCCCACTTCCGCTCACTCTCCCACTCC 1925

Qy     521 CysSerIleValThrLysThrTyrGlnGlyTrpGluGlyValThrAsnValAsnThrPro 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1926 TGCAGTATCGTCACTAAGACCTACCAGGGTTGGGAAGGCGTCACGAACGTCAACACGCCT 1985

Qy     541 PheGlyGlnPheAlaHisAlaGlyLeuLeuLysAsnGluGluIleLeuCysLeuAlaAsp 560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1986 TTCGGCCAATTCGCGCACGCGGGCCTCCTCAAGAATGAGGAGATCCTCTGCCTCGCCGAC 2045

Qy     561 AspLeuAlaThrArgLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaAlaVal 580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2046 GACCTGGCCACCCGTCTCACAGGTGTCTACCCCGCCACTGACAACTTCGCGGCCGCCGTT 2105

Qy     581 SerAlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGluAlaThrSerSerIle 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2106 TCTGCCTTCGCCGCGAACATGCTGTCCTCCGTGCTGAAGTCGGAGGCAACGTCCTCCATC 2165

Qy     601 IleLysSerValGlyGluThrAlaValGlyAlaAlaGlnSerGlyLeuAlaLysLeuPro 620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2166 ATCAAGTCCGTTGGCGAGACTGCCGTCGGCGCGGCTCAGTCCGGCCTCGCGAAGCTACCC 2225

Qy     621 GlyLeuLeuMetSerValProGlyLysIleAlaAlaArgValArgAlaArgArgAlaArg 640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2226 GGACTGCTAATGAGTGTACCAGGGAAGATTGCCGCGCGTGTCCGCGCGCGCCGAGCGCGC 2285

Qy     641 ArgArgAlaAlaArgAlaAsn 647
           |||||||||||||||||||||
Db    2286 CGCCGCGCCGCTCGTGCCAAT 2306
```

RESULT 2
AAQ58523
ID AAQ58523 standard; DNA; 2478 BP.
XX
AC AAQ58523;
XX
DT 12-SEP-1994 (first entry)
XX
DE Sequence of Heliothis armigera RNA 2 which encodes P17 and the
DE capsid proteins precursor P71.
XX
KW HaSV; RNA 1; small RNA virus; P17; P71; ss.
XX
OS Heliothis armigera stunt virus.
XX
FH Key   Location/Qualifiers
FT CDS   283..753
FT   /*tag= a
FT   /product= P17
FT CDS   366..2309
FT   /*tag= b
FT   /product= P71
XX
PN WO9404660-A.
XX
PD 03-MAR-1994.
XX
PF 13-AUG-1993; 93WO-AU00411.
XX
PR 14-AUG-1992; 92AU-0004081.
PR 08-JUL-1993; 93US-0089372.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (PACI-) PACIFIC SEEDS PTY LTD.
XX
PI Christian PD, Gordon KHJ, Hanzlik TN;
XX
DR WPI; 1994-083180/10.
DR P-PSDB; AAR49661, AAR49662.
XX
PT Small RNA virus capable of infecting insect species, e.g.
PT Heliothis - and transgenic plants contg. viral nucleic acid, for
PT protection against insect pests
XX
PS Disclosure; Figure 2; 183pp; English.
XX
CC The inventors claim a virus comprising a genome hybridisable with
CC the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are
CC those given in Figs 1 and 2 of the specification. As isolated
CC protein or polypeptide prepn. of the proteins or polypeptides
CC derivable from the virus are also claimed.
CC H. armigera larvae were raised and viral RNA was extracted. The virus
CC RNAs were reverse transcribed into cDNA. Clone hr236 contains about
CC 88% or RNA 2. RNA 2 encodes a protein of mol.wt.71,000(P71) which
CC contains the peptide sequences corresp. to those determined from
CC the two virus capsid proteins. This protein is therefore the
CC precursor of these capsid proteins. In addition, another major
CC translation product of apparent mol. wt. 24,000 is obtd. This
CC protein is derived from a mol. wt. 17,000 reading frame overlapping
CC the slab of the capsid protein gene. The Mr 24,000 protein (referred
CC to as P17) may have a function in modifying or manipulating the growth
CC characteristics or cell cycle of HaSV-infected cells.
XX
SQ Sequence 2478 BP; 516 A; 853 C; 608 G; 501 T; 0 other;

Alignment Scores:

Pred. No.:	1.92e-272	Length:	2478
Score:	3353.00	Matches:	644
Percent Similarity:	99.54%	Conservative:	0
Best Local Similarity:	99.54%	Mismatches:	3
Query Match:	99.38%	Indels:	0
DB:	15	Gaps:	0

US-09-677-653A-50 (1-647) x AAQ58523 (1-2478)

```
Qy       1 MetGlyAspAlaGlyValAlaSerGlnArgProHisAsnArgArgGlyThrArgAsnVal 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 ATGGGAGATGCTGGAGTGGCGTCACAGCGACCTCACAACCGTCGCGGAACCCGTAACGTT 425

Qy      21 ArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArgThrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 CGGGTCAGCGCCAACACCGTCACCGTCAATGGTAGAAGAAACCAACGGCGTCGGACCGGA 485

Qy      41 ArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeu 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     486 AGGCAAGTTTCTCCCCCTGACAATTTCACCGCTGCTGCACAAGACCTCGCGCAAAGCCTT 545

Qy      61 AspAlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrp 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     546 GACGCCAACACCGTCACTTTCCCCGCTAACATCTCTAGCATGCCCGAATTCCGGAATTGG 605

Qy      81 AlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTrpTyrPheLysTyrLeuAsp 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     606 GCCAAGGGAAAGATCGACCTCGACTCCGATTCCATCGGCTGGTACTTCAAGTACCTTGAC 665

Qy     101 ProAlaGlyAlaThrGluSerAlaArgAlaValGlyGluTyrSerLysIleProAspGly 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     666 CCAGCGGGTGCTACAGAGTCTGCGCGCGCCGTCGGCGAGTACTCGAAGATCCCTGACGGC 725

Qy     121 LeuValLysPheSerValAspAlaGluIleArgGluIleTyrAsnGluGluCysProVal 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     726 CTCGTCAAGTTCTCCGTCGACGCAGAGATAAGAGAGATCTATAACGAGGAGTGCCCCGTC 785

Qy     141 ValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPhe 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     786 GTCACTGACGTGTCCGTCCCCCTCGACGGCCGCCAGTGGAGCCTCTCGATTTTCTCCTTT 845

Qy     161 ProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMetSerLeu 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 CGACGTTTCAGAACCGCCTACGTCGCCGTAGCGAACGTCGAGAACAAGGAGATGTCGCTC 905

Qy     181 AspValValAsnAspLeuIleGluTrpLeuAsnAsnLeuAlaAspTrpArgTyrValVal 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    906 GACGTTGTCAACGACCTCATCGAGTGGCTCAACAATCTCGCCGACTGGCGTTATGTCGTT 965

Qy    201 AspSerGluGlnTrpIleAsnPheThrAsnAspThrThrTyrTyrValArgIleArgVal 220
          ||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db    966 GACTCTGAACAGTGGATTAACTTCACCAATGACACCAGGTACTACGTCCGCATCCGCGTT 1025

Qy    221 LeuArgProThrTyrAspValProAspProThrGluGlyLeuValArgThrValSerAsp 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 CTACGTCCAACCTACGACGTTCCAGACCCCACAGAGGGCCTTGTTCGCACAGTCTCAGAC 1085

Qy    241 TyrArgLeuThrTyrLysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGln 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1086 TACCGCCTCACTTATAAGGCGATAACATGTGAAGCCAACATGCCAACACTCGTCGACCAA 1145

Qy    261 GlyPheTrpIleGlyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspVal 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1146 GGCTTTTGGATCGGCGGCCAGTACGCTCTCACCCCGACTAGCCTACCGCAGTACGACGTC 1205

Qy    281 SerGluAlaTyrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1206 AGCGAGGCCTACGCTCTGCACACTTTGACCTTCGCCAGACCATCCAGCGCCGCTGCACTC 1265

Qy    301 AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrp 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1266 GCGTTTGTGTGGGCAGGTTTGCCACAGGGTGGCACTGCGCCTGCAGGCACTCCAGCCTGG 1325

Qy    321 GluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPheProAla 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1326 GAGCAGGCATCCTCGGGTGGCTACCTCACCTGGCGCCACAACGGTACTACTTTCCCAGCT 1385

Qy    341 GlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspProAsnAsp 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1386 GGCTCCGTTAGCTACGTTCTCCCTGAGGGTTTCGCCCTTGAGCGCTACGACCCGAACGAC 1445

Qy    361 GlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaVal 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1446 GGCTCTTGGACCGACTTCGCTTCCGCAGGAGACACCGTCACTTTCCGGCAGGTCGCCGTC 1505

Qy    381 AspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPheThrVal 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1506 GACGAGGTCGTTGTGACCAACAACCCCGCCGGCGGCGGCAGCGCCCCCACCTTCACCGTG 1565

Qy    401 ArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeuGluThr 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1566 AGAGTGCCCCCTTCAAACGCTTACACCAACACCGTGTTTAGGAACACGCTCTTAGAGACT 1625

Qy    421 ArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrVal 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1626 CGACCCTCCTCTCGTAGGCTCGAACTCCCTATGCCACCTGCTGACTTTGGACAGACGGTC 1685

Qy    441 AlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeuGlyCysTyrLeuVal 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1686 GCCAACAACCCGAAGATCGAGCAGTCGCTTCTTAAAGAAACACTTGGCTGCTATTTGGTC 1745

Qy    461 HisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaVal 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1746 CACTCCAAAATGCGAAACCCCGTTTTCCAGCTCACGCCAGCCAGCTCCTTTGGCGCCGTT 1805

Qy    481 SerPheAsnAsnProGlyTyrGluArgThrArgAspLeuProAspTyrThrGlyIleArg 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1806 TCCTTCAACAATCCGGGTTATGAGCGCACACGCGACCTCCCGGACTACACTGGCATCCGT 1865

Qy    501 AspSerPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1866 GACTCATTCGACCAGAACATGTCCACCGCTGTGGCCCACTTCCGCTCACTCTCCCACTCC 1925

Qy    521 CysSerIleValThrLysThrTyrGlnGlyTrpGluGlyValThrAsnValAsnThrPro 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1926 TGCAGTATCGTCACTAAGACCTACCAGGGTTGGGAAGGCGTCACGAACGTCAACACGCCT 1985

Qy    541 PheGlyGlnPheAlaHisAlaGlyLeuLeuLysAsnGluGluIleLeuCysLeuAlaAsp 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1986 TTCGGCCAATTCGCGCACGCGGGCCTCCTCAAGAATGAGGAGATCCTCTGCCTCGCCGAC 2045

Qy    561 AspLeuAlaThrArgLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaAlaVal 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2046 GACCTGGCCACCCGTCTCACAGGTGTCTACCCCGCCACTGACAACTTCGCGGCCGCCGTT 2105

Qy    581 SerAlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGluAlaThrSerSerIle 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2106 TCTGCCTTCGCCGCGAACATGCTGTCCTCCGTGCTGAAGTCGGAGGCAACGTCCTCCATC 2165

Qy    601 IleLysSerValGlyGluThrAlaValGlyAlaAlaGlnSerGlyLeuAlaLysLeuPro 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2166 ATCAAGTCCGTTGGCGAGACTGCCGTCGGCGCGGCTCAGTCCGGCCTCGCGAAGCTACCC 2225

Qy    621 GlyLeuLeuMetSerValProGlyLysIleAlaAlaArgValArgAlaArgArgAlaArg 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2226 GGACTGCTAATGAGTGTACCAGGGAAGATTGCCGCGCGTGTCCGCGCGCGCCGAGCGCGC 2285

Qy    641 ArgArgAlaAlaArgAlaAsn 647
          ||||||||||||||||||||
Db   2286 CGCCGCGCCGCTCGTGCCAAT 2306
```

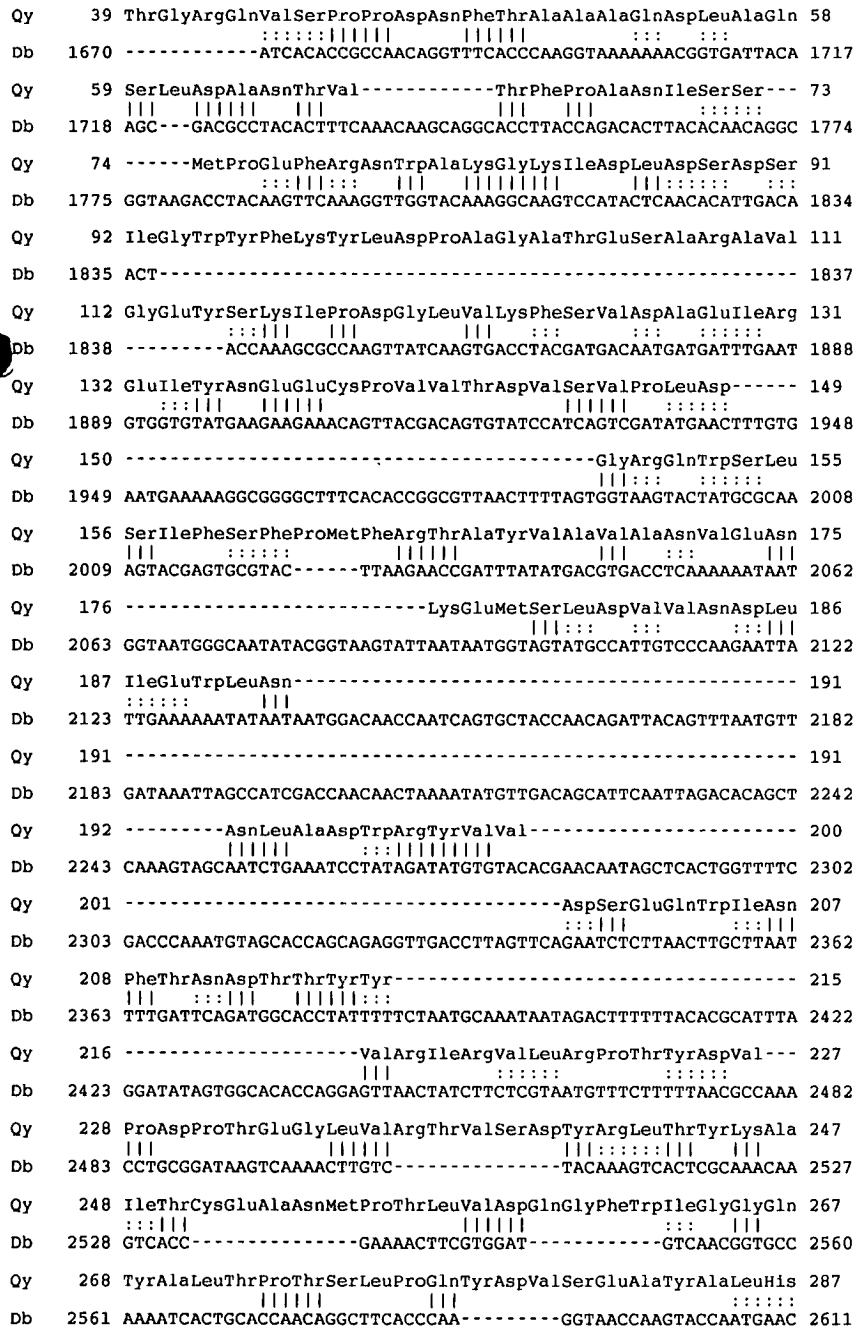
```
RESULT 3
AAV04471
ID   AAV04471 standard; cDNA; 6536 BP.
XX
AC   AAV04471;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Nudaurelia beta-like virus RNA genome cDNA.
XX
KW   NBV; RNA virus; transgenic plant; insect resistance;
KW   disease resistance; Nadaurelia cytherea capensis;
KW   pine tree emperor moth; virus-like particle; vector; replicase;
KW   capsid protein; ds.
XX
OS   Nadaurelia beta-like virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             31..3732
FT                   /*tag= a
FT                   /product= replicase
FT   CDS             3884..5788
FT                   /*tag= b
FT                   /product= capsid protein precursor
FT   CDS             4219..5112
FT                   /*tag= c
FT                   /note= "encodes protein of unknown function"
XX
PN   AU9724669-A.
XX
PD   04-DEC-1997.
XX
PF   02-JUN-1997;   97AU-0024669.
XX
PR   31-MAY-1996;   96AU-0000233.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (UYRH-) UNIV RHODES.
XX
PI   Gordon KH,  Hanzlik TN,  Hendry DA;
XX
DR   WPI; 1998-052736/06.
DR   P-PSDB; AAW41935 AND AAW26785.
XX
PT   Nudaurelia beta virus nucleic acid - useful for producing
PT   recombinant virus, insect-resistant transgenic plants, etc
XX
PS   Claim 3; Fig 1; 33pp; English.
XX
CC   This cDNA corresponds to the RNA genome of Nadaurelia beta-like
CC   virus (NBV).  cDNA fragments were generated by random-primed
CC   reverse transcription of RNA of NBV collected from Nadaurelia
```

```
CC   cytherea capensis larvae.  These were cloned into pBluescript II
CC   KS(-), and inserts were sequenced to identify the 6536 bp sequence.
CC   Sequence analysis indicated that the RNA genome is dicistronic,
CC   with a 5' open reading frame (ORF) encoding a replicase (see
CC   AAW41935) and a 3' ORF encoding a 70 kDa capsid protein precursor (see
CC   AAW26785).  A thrid, overlapping, ORF codes for a protein of unknown
CC   function.  A claimed infectious recombinant insect virus vector
CC   comprises an expressible nucleic acid molecule comprising a
CC   nucleotide sequence corresponding to all or an infectious and/or
CC   insecticidal portion of the genomic RNA of NBV.  Also claimed is: a
CC   transgenic plant resistant to insect attack that produces NBV such
CC   that insects feeding on the plant are deleteriously affected; and a
CC   virus-like particle (VLP) prepared from expression of a nucleic acid
CC   molecule comprising a sequence encoding the capsid protein of NBV.
CC   The invention provides methods for the control of insect pests (the
CC   pine-tree emperor moth Nudaurelia cytherea capensis is mentioned).
XX
SQ   Sequence 6536 BP; 1678 A; 1770 C; 1751 G; 1336 T; 1 other;

Alignment Scores:
Pred. No.:                2.82e-14        Length:          6536
Score:                    294.00          Matches:         184
Percent Similarity:       37.50%          Conservative:    74
Best Local Similarity:    26.74%          Mismatches:      281
Query Match:              8.71%           Indels:          151
DB:                       19              Gaps:            35

US-09-677-653A-50 (1-647) x AAV04471 (1-6536)

Qy      2 GlyAspAlaGlyValAlaSerGlnArgProHisAsnArgArgGlyThrArgAsnValArg 21
          |||      |||         |||:::         ||||||      |||   ::::::
Db   4014 GGGCAACCAGGGCCGCGCAAACAACAACAACAACGTCGACGCAGGAGGCGGGGTCTAAAA 4073

Qy     22 ValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArgThrGlyArg 41
          :::                  :::
Db   4074 CTTCCCCCCGTAGTCGCACCGATT---------------------------------- 4097

Qy     42 GlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeuAsp 61
                |||||||||||      |||
Db   4098 ---ACCAGCCCGCCAGATGGCCGAACCCCGAAC----------------------CAC 4130

Qy     62 AlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrpAla 81
          |||||||||        |||   :::              :::   |||      |||
Db   4131 GCGAACACCCGGGTCAA-CCGCGGTCG--------------GACACGCGTGAGGGGGCT 4174

Qy     82 LysGlyLysIleAspLeuAsp---------SerAspSerIleGlyTrpTyrPheLysTyr 98
          |||         |||   :::         ||||||         |||         |||
Db   4175 AAGGCAAGCGATGATGGAGAGTCCCATGGCAGCGACATCAAGGCATGGATTCACGACTAT 4234

Qy     99 LeuAspProAlaGlyAlaThrGluSerAlaArgAlaValGlyGluTyrSerLysIlePro 118
          |||||||||   |||      :::::::::          |||         |||||||||
Db   4235 CTAGACCCGGACGGAGAATACAAGACGAGCCTGGACGACGGG---------AAAATTCCC 4285

Qy    119 AspGlyLeuValLysPheSerValAspAlaGluIleArgGluIleTyrAsnGluGluCys 138
          ||||||   :::       |||         :::   |||
Db   4286 GACGGCGCGATACCTCAGTCAACATGCGGTCAATTTCGAGGGACCGTGGGCGCCAGATAC 4345

Qy    139 ProValValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePhe 158
          |||   :::         ::::::||||||||||||      |||   |||   :::
Db   4346 CCGGGACTGAATTCTACGACGCTACCGCTGGATGGCGGGACCTGGCCTCTACTAGTGATG 4405

Qy    159 SerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMet 178
             |||   ||||||        :::   :::          |||   |||
Db   4406 CATCTCCCGTTCTTCAGGCATCCGTTGTTGTTCATCACCACCACCAGCAACACGGAA--- 4462

Qy    179 SerLeuAspValValAsn---AspLeu------------IleGluTrpLeuAsnAsnLeu 193
             ::::::|||   |||   ||||||              :::|||   ||||||
Db   4463 ---GTCGAAGTGACGAACGCCGATCTGGATGCGTTCGCGAACGATTGG---AACAACAGG 4516

Qy    194 AlaAspTrpArgTyrValValAspSerGluGlnTrpIleAsnPheThrAsnAspThrThr 213
             ||||||                        |||         |||
Db   4517 ACGGACTGGACCGAAGCGACG---TACCCAAGTTGGGCGCAAGTCGGGAAC------GTG 4567

Qy    214 TyrTyrValArgIleArgValLeuArgProThr---------TyrAspValProAspPro 230
          :::|||:::   :::           ||||||          |||||||||   |||
Db   4568 TTTTACATGGTCGTC-----------CCGACCGAAGCGCTGACGGACGTACCACCCCCG 4615

Qy    231 ThrGlu---GlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAlaIleThr 249
          |||:::   |||:::         :::      ||||||||||||         :::|||
Db   4616 ACTCAACTGGGTGTATCAGGGTTACTCGAGAGTTACCGTCTGACATCGAGCGGCGTCACA 4675

Qy    250 CysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGlnTyrAla 269
                |||   ||||||||||||:::||||||      :::      |||:::
Db   4676 GCGTACTTCAACGCACCCACTCTCGTGAATCAGGGAGTGGCGGTGATCGCGCAGTTC--- 4732

Qy    270 LeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHisThrLeu 289
             |||         |||  :::      :::      |||      |||
Db   4733 ---CAACCGGACAAAGAACACCAGAAGGAGAACCCGGACATAGTAGCCGGGACCACCCAA 4789

Qy    290 Thr---------------------PheAlaArgProSerSerAlaAlaAlaLeuAlaPhe 302
          |||                        :::   |||:::           :::
Db   4790 ACGGGCGGAACGTTACAGCTCGGCGGTTCAGGGCCGAACTACACATTGACGATGACG--- 4846

Qy    303 ValTrpAlaGlyLeuProGlnGlyGlyThrAla---ProAlaGlyThrProAlaTrpGlu 321
          :::         :::      ||||||   |||   |||      |||   :::
Db   4847 ATCGGGGACCAGGTCGAGTTCGGGGGCGCAGCAATCCCGCTACCCACGGTGTCGATGGGG 4906

Qy    322 GlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrProProAlaGly 341
                   |||   |||  ::::::              ||||||      |||
Db   4907 CCGATGCCGGAGTCGGGGCAGCTGGTGTTCCAGACTGCGAACCTGACATTCGACGTCGGA 4966

Qy    342 ---------SerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358
                   :::::::::      ||||||   |||
Db   4967 AACACAATCACCATCACGACCACGCTGCCACCAGGGTCGGTG----------------- 5008

Qy    359 AsnAspGlySerTrpThrAspPheAlaSerAlaGly---AspThrValThrPheArgGln 377
             |||   |||         ||||||   |||   ||||||||||||
Db   5009 ---ACGGGAATGTGGCAATTCACAGCCAGCAACGGGACGGACACCGTGACC--------- 5056

Qy    378 ValAlaValAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThr 397
             ||||||                        ||||||
Db   5057 ------GTGGAC-----------------------GCGGGAGCG--------------- 5071

Qy    398 PheThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeu 417
          ||||||||||||                        |||
Db   5072 ---ACTGTACGCGTTCGGAGCGAATTTGGACGCCTCGGAACTGAATCTGCAGGACATCAA 5128

Qy    418 LeuGluThrArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGly 437
          |||             ::::::                  |||   |||   |||   |||
Db   5129 CTCAATCAAGATTCCACCAAC--------------GACATGAACCCAAATGATGCAGGC 5173

Qy    438 GlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLys-GluThrLeuGlyCy 457
                   |||         |||:::      |||   ||| |||   :::   ||
Db   5174 -----------AACGCCAAGACCATTCAGTTCCAACTAACGAAACGAGGGCATTATATG 5221

Qy    457 sTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPh 477
          |                  |||      ||||||::::::|||   |||:::|||::
Db   5222 CCTG--------------AGGCGTCAATCCGTGTTCGAAATGACAAT-GCGACGTCTTA 5265

Qy    477 eGlyAlaValSerPheAsnAsnProGlyTyrGluArgThrArgAsp----LeuProAspT 496
          :|||   |||                      |||   |||   |||    :::   ||||
Db   5266 TGGACCGGTCGAT-----------------GAAGACACCGAGGACAACTGTGGTAGATT 5307

Qy    496 yrThrGly------------IleArgAspSerPheAspGlnAsnMetSer---------- 508
          |||||||              :::|||:::   |||   |||   :::
Db   5308 ACACCGGGCAATTGGTGCACTCCAAGGATACCATCGACAGCAACTTCGCGATAGGTTGCC 5367

Qy    509 --------ThrAlaValAlaHis-PheArgSerLeuSerHisSerCysSerIleValThr 525
                            |||    |||:::|||         ||||||      :::
Db   5368 GTCGATGACCGGTATGTCTACATCAACCGTACCCTACTTGCAAGGTGTTCCGACGCTTCG 5427

Qy    526 LysThrTyrGln-GlyTrpGluGlyValThrAsnValAsnThrProPheGlyGlnPheAl 545
          |||   |||:: |||   |||                      |||:::|||   |||||
Db   5428 AAGCGATACCGGCGGAGGGGAG--------------------CCTTGGGGCCCCTTCGC 5466

Qy    545 aHisAlaGlyLeuLeuLysAsnGluGluIleLeuCysLeuAlaAspAspLeuAlaThrAr 565
```

```
Db   5467 TAGTGCGACACCTCCGAAGGACGACGTGGCGCTAACAGTGGCTCGAACTTGGACCGATCT 5526

Qy    565 gLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaAlaValSerAlaPheAlaAl 585

Db   5527 GCACCCATTCGCATACCCGGAACGATACAACGGATTCGGGGCCCTATTCGCGATGGTGGC 5586

Qy    585 aAsnMetLeuSerSerValLeuLysSerGluAlaThrSerSerIleIleLysSerValGl 605

Db   5587 CAAGACCATAGCCCAGATACCTCGC----------------TATGTGCGATCA----- 5623

Qy    605 yGluThrAlaValGlyAlaAlaGlnSerGlyLeuAlaLysLeuProGlyLeuLeuMetSe 625

Db   5624 -------GCAGCCGGAGTGGCGAATGCGGTGACGGACTGCATAGAGAGCGCGACCGAGAG 5676

Qy    625 rValProGlyLysIleAlaAlaArgValArgAlaArgArgAlaArgArg---------- 641

Db   5677 TGTAGCCTCGAATTCCACCTCGGAGAGGCGGCAACGAAGAGCGAGACGTGTTGGCGGAAT 5736

 y    642 ----ArgAlaAlaArg 645

Db   5737 CGCTCGAGGAGCCCGC 5752

RESULT 4
AAT99118
ID   AAT99118 standard; cDNA; 6536 BP.
XX
AC   AAT99118;
XX
DT   08-JUN-1998  (first entry)
XX
DE   Nudaurelia beta virus RNA genome encoding coat protein p70.
XX
KW   Vaccine; coat protein; p70; insecticide; Ig-like domain; ds.
XX
OS   Nudaurelia beta virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             31..3732
FT                   /*tag= a
FT                   /product= undefined protein
FT   CDS             3884..5788
FT                   /*tag= b
FT                   /product= p70 coat protein
XX
PN   WO9746666-A1.
XX
PD   11-DEC-1997.
XX
PF   02-JUN-1997;   97WO-AU00349.
XX
PR   31-MAY-1996;   96AU-0000234.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Gordon KH,  Hanzlik TN;
XX
DR   WPI; 1998-042175/04.
DR   P-PSDB; AAW34536, AAW34537.
XX
PT   Modified small RNA viruses and virus-like particles - have altered
PT   or substituted Ig-like domains to modify host cell tropism, useful
PT   as insecticides and in medicinal applications
XX
PS   Disclosure; Figure 2; 41pp; English.
XX
CC   The sequence is that of a cDNA encoding the viral genome RNA which
CC   putatively codes for coat protein p70. The p70 coat protein
CC   includes an Ig-like domain which can be used in the production of
CC   virus-like particles (VLPs). The VLPs can be used in vaccines where the
CC   Ig-like domain has been altered so that the VLP presents a surface
CC   located antigen which is used to elicit an immune response in a host
CC   organism. They can also be used controlling the proliferation of a pest
CC   insect and potentially as medicinal delivery agents for cancer treatment
CC   and gene therapy.
XX
SQ   Sequence 6536 BP; 1679 A; 1770 C; 1751 G; 1335 T; 1 other;

Alignment Scores:
Pred. No.:              2.82e-14        Length:         6536
Score:                  294.00          Matches:        184
Percent Similarity:     37.50%          Conservative:   74
Best Local Similarity:  26.74%          Mismatches:     281
Query Match:            8.71%           Indels:         151
DB:                     19              Gaps:           35

US-09-677-653A-50 (1-647) x AAT99118 (1-6536)

Qy      2 GlyAspAlaGlyValAlaSerGlnArgProHisAsnArgArgGlyThrArgAsnValArg 21

Db   4014 GGGCAACCAGGGCCGCGCAAACAACAACAACAACGTCGACGCAGGAGGCGGGGTCTAAAA 4073

Qy     22 ValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArgThrGlyArg 41

Db   4074 CTTCCCCCCGTAGTCGCACCGATT---------------------------------- 4097

Qy     42 GlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeuAsp 61

Db   4098 ---ACCAGCCCGCCAGATGGCCGAACCCCGAAC----------------------CAC 4130

Qy     62 AlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrpAla 81

Db   4131 GCGAACACCCGGGTCAA-CCGCGGTCG--------------GACACGCGTGAGGGGGCT 4174

Qy     82 LysGlyLysIleAspLeuAsp---------SerAspSerIleGlyTrpTyrPheLysTyr 98

Db   4175 AAGGCAAGCGATGATGGAGAGTCCCATGGCAGCGACATCAAGGCATGGATTCACGACTAT 4234

Qy     99 LeuAspProAlaGlyAlaThrGluSerAlaArgAlaValGlyGluTyrSerLysIlePro 118

Db   4235 CTAGACCCGGACGGAGAATACAAGACGAGCCTGGACGACGGG---------AAAATTCCC 4285

Qy    119 AspGlyLeuValLysPheSerValAspAlaGluIleArgGluIleTyrAsnGluGluCys 138

Db   4286 GACGGCGCGATACCTCAGTCAACATGCGGTCAATTTCGAGGGACCGTGGGCGCCAGATAC 4345

Qy    139 ProValValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePhe 158

Db   4346 CCGGGACTGAATTCTACGACGCTACCGCTGGATGGCGGGACCTGGCCTCTACTAGTGATG 4405

Qy    159 SerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMet 178

Db   4406 CATCTCCCGTTCTTCAGGCATCCGTTGTTGTTCATCACCACCACCAGCAACACGGAA--- 4462

Qy    179 SerLeuAspValValAsn---AspLeu------------IleGluTrpLeuAsnAsnLeu 193

Db   4463 ---GTCGAAGTGACGAACGCCGATCTGGATGCGTTCGCGAACGATTGG---AACAACAGG 4516

Qy    194 AlaAspTrpArgTyrValValAspSerGluGlnTrpIleAsnPheThrAsnAspThrThr 213

Db   4517 ACGGACTGGACCGAAGCGACG---TACCCAAGTTGGGCGCAAGTCGGGAAC------GTG 4567

Qy    214 TyrTyrValArgIleArgValLeuArgProThr---------TyrAspValProAspPro 230

Db   4568 TTTTACATGGTCGTC------------CCGACCGAAGCGCTGACGGACGTACCACCCCCG 4615

Qy    231 ThrGlu---GlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAlaIleThr 249

Db   4616 ACTCAACTGGGTGTATCAGGGTTACTCGAGAGTTACCGTCTGACATCGAGCGGCGTCACA 4675

Qy    250 CysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGlnTyrAla 269

Db   4676 GCGTACTTCAACGCACCCACTCTCGTGAATCAGGGAGTGGCGGTGATCGCGCAGTTC--- 4732

Qy    270 LeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHisThrLeu 289
```

```
Db   4733 ---CAACCGGACAAAGAACACCAGAAGGAGAACCCGGACATAGTAGCCGGGACCACCCAA 4789

Qy    290 Thr--------------------PheAlaArgProSerSerAlaAlaAlaLeuAlaPhe 302
          |||                        :::   |||:::           :::
Db   4790 ACGGGCGGAACGTTACAGCTCGGCGGTTCAGGGCCGAACTACACATTGACGATGACG--- 4846

Qy    303 ValTrpAlaGlyLeuProGlnGlyGlyThrAla---ProAlaGlyThrProAlaTrpGlu 321
          :::         :::      ||||||   |||   |||      |||   :::
Db   4847 ATCGGGGACCAGGTCGAGTTCGGGGGCGCAGCAATCCCGCTACCCACGGTGTCGATGGGG 4906

Qy    322 GlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrProAlaGly 341
                |||  |||   ::::::        ||||||      |||
Db   4907 CCGATGCCGGAGTCGGGGCAGCTGGTGTTCCAGACTGCGAACCTGACATTCGACGTCGGA 4966

Qy    342 ---------SerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358
                  :::::::::      ||||||   |||
Db   4967 AACACAATCACCATCACGACCACGCTGCCACCAGGGTCGGTG----------------- 5008

Qy    359 AsnAspGlySerTrpThrAspPheAlaSerAlaGly---AspThrValThrPheArgGln 377
          |||  |||         ||||||   |||   ||||||||||||
Db   5009 ---ACGGGAATGTGGCAATTCACAGCCAGCAACGGGACGGACACCGTGACC--------- 5056

Qy    378 ValAlaValAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThr 397
          ||||||                      ||||||
Db   5057 ------GTGGAC----------------------GCGGGAGCG-------------- 5071

Qy    398 PheThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeu 417
             ||||||||||||                    |||
Db   5072 ---ACTGTACGCGTTCGGAGCGAATTTGGACGCCTCGGAACTGAATCTGCAGGACATCAA 5128

Qy    418 LeuGluThrArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGly 437
          |||           ::::::                |||   |||   |||   |||
Db   5129 CTCAATCAAGATTCCACCAAC--------------GACATGAACCCAAATGATGCAGGC 5173

Qy    438 GlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLys-GluThrLeuGlyCy 457
                |||         |||:::      |||   ||| |||   :::   ||
Db   5174 -----------AACGCCAAGACCATTCAGTTCCAACTAACGAAACGAGGGCATTATATG 5221

Qy    457 sTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPh 477
          |                  |||      ||||||::::::|||   |||:::|||::
Db   5222 CCTG--------------AGGCGTCAATCCGTGTTCGAAATGACAAT-GCGACGTCTTA 5265

Qy    477 eGlyAlaValSerPheAsnAsnProGlyTyrGluArgThrArgAsp----LeuProAspT 496
          :|||   |||                    |||   |||   |||    :::   ||||
Db   5266 TGGACCGGTCGAT----------------GAAGACACCGAGGACAACTGTGGTAGATT 5307

Qy    496 yrThrGly------------IleArgAspSerPheAspGlnAsnMetSer---------- 508
          |||||||             :::|||:::   |||   |||   :::
Db   5308 ACACCGGGCAATTGGTGCACTCCAAGGATACCATCGACAGCAACTTCGCGATAGGTTGCC 5367

Qy    509 --------ThrAlaValAlaHis-PheArgSerLeuSerHisSerCysSerIleValThr 525
                      |||    |||:::|||         ||||||      :::
Db   5368 GTCGATGACCGGTATGTCTACATCAACCGTACCCTACTTGCAAGGTGTTCCGACGCTTCG 5427

Qy    526 LysThrTyrGln-GlyTrpGluGlyValThrAsnValAsnThrProPheGlyGlnPheAl 545
          |||   |||::: |||   |||                  |||:::|||   |||||
Db   5428 AAGCGATACCGGCGGAGGGGAG-------------------CCTTGGGGCCCCTTCGC 5466

Qy    545 aHisAlaGlyLeuLeuLysAsnGluGluIleLeuCysLeuAlaAspAspLeuAlaThrAr 565
          |   |||         |||::::::      |||   :::|||
Db   5467 TAGTGCGACACCTCCGAAGGACGACGTGGCGCTAACAGTGGCTCGAACTTGGACCGATCT 5526

Qy    565 gLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaAlaValSerAlaPheAlaAl 585
                |||||         |||         |||:::   |||      ||
Db   5527 GCACCCATTCGCATACCCGGAACGATACAACGGATTCGGGGCCCTATTCGCGATGGTGGC 5586

Qy    585 aAsnMetLeuSerSerValLeuLysSerGluAlaThrSerSerIleIleLysSerValGl 605
          |      ::::::   :::   :::                  ::::::|||
Db   5587 CAAGACCATAGCCCAGATACCTCGC-----------------TATGTGCGATCA----- 5623

Qy    605 yGluThrAlaValGlyAlaAlaGlnSerGlyLeuAlaLysLeuProGlyLeuLeuMetSe 625
          |||   |||   |||   :::           :::                  ||
Db   5624 -------GCAGCCGGAGTGGCGAATGCGGTGACGGACTGCATAGAGAGCGCGACCGAGAG 5676
```

```
Qy    625 rValProGlyLysIleAlaAlaArgValArgAlaArgArgAlaArgArg---------- 641
          ||||              :::      |||   ||||||||||||||
Db   5677 TGTAGCCTCGAATTCCACCTCGGAGAGGCGGCAACGAAGAGCGAGACGTGTTGGCGGAAT 5736

Qy    642 ----ArgAlaAlaArg 645
                |||   ||||||
Db   5737 CGCTCGAGGAGCCCGC 5752

RESULT 5
AAX20209
ID   AAX20209 standard; DNA; 3695 BP.
XX
AC   AAX20209;
XX
DT   20-APR-1999  (first entry)
XX
DE   Enterococcus faecalis EF108 gene fragment.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic; ss.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C,  Choi GH,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   P-PSDB; AAY00219.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 1; Page 210-211; 301pp; English.
XX
CC   The present sequence encodes an antigenic polypeptide fragment
CC   isolated from Enterococcus faecalis. The present invention describes
CC   genes, proteins and antigenic polypeptides isolated from E. faecalis.
CC   The proteins can be used in vaccines for preventing or attenuating an
CC   infection caused by a member of the Enterococcus genus in an animal.
CC   They can also be used for detecting Enterococcus antibodies in a sample.
CC   The nucleotide sequences can be used for detecting Enterococcus nucleic
CC   acids. Products from the present invention can also be used for
CC   screening compounds to identify agonists and antagonists of E. faecalis
CC   protein activity.
XX
SQ   Sequence 3695 BP; 1293 A; 790 C; 739 G; 873 T; 0 other;

Alignment Scores:
Pred. No.:              0.131          Length:         3695
Score:                  139.50         Matches:        140
Percent Similarity:     31.38%         Conservative:   95
Best Local Similarity:  18.69%         Mismatches:     247
Query Match:            4.13%          Indels:         267
DB:                     20             Gaps:           31

US-09-677-653A-50 (1-647) x AAX20209 (1-3695)

Qy     19 AsnValArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArg 38
          |||   :::|||:::   |||   |||      ||||||   :::
Db   1625 AACCGCAAAGTCACCGAGAACTTCGTAGACACCAACGGCGCTAAA--------------- 1669
```


Qy 39 ThrGlyArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGln 58
Db 1670 -----------ATCACACCGCCAACAGGTTTCACCCAAGGTAAAAAAACGGTGATTACA 1717
Qy 59 SerLeuAspAlaAsnThrVal-----------ThrPheProAlaAsnIleSerSer--- 73
Db 1718 AGC---GACGCCTACACTTTCAAACAAGCAGGCACCTTACCAGACACTTACACAACAGGC 1774
Qy 74 ------MetProGluPheArgAsnTrpAlaLysGlyLysIleAspLeuAspSerAspSer 91
Db 1775 GGTAAGACCTACAAGTTCAAAGGTTGGTACAAAGGCAAGTCCATACTCAACACATTGACA 1834
Qy 92 IleGlyTrpTyrPheLysTyrLeuAspProAlaGlyAlaThrGluSerAlaArgAlaVal 111
Db 1835 ACT---------- 1837
Qy 112 GlyGluTyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArg 131
Db 1838 ---------ACCAAAGCGCCAAGTTATCAAGTGACCTACGATGACAATGATGATTTGAAT 1888
Qy 132 GluIleTyrAsnGluGluCysProValValThrAspValSerValProLeuAsp------ 149
Db 1889 GTGGTGTATGAAGAAGAAACAGTTACGACAGTGTATCCATCAGTCGATATGAACTTTGTG 1948
Qy 150 ----------GlyArgGlnTrpSerLeu 155
Db 1949 AATGAAAAAGGCGGGGCTTTCACACCGGCGTTAACTTTTAGTGGTAAGTACTATGCGCAA 2008
Qy 156 SerIlePheSerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsn 175
Db 2009 AGTACGAGTGCGTAC------TTAAGAACCGATTTATATGACGTGACCTCAAAAAATAAT 2062
Qy 176 ----------LysGluMetSerLeuAspValValAsnAspLeu 186
Db 2063 GGTAATGGGCAATATACGGTAAGTATTAATAATGGTAGTATGCCATTGTCCCAAGAATTA 2122
Qy 187 IleGluTrpLeuAsn---------- 191
Db 2123 TTGAAAAAATATAATAATGGACAACCAATCAGTGCTACCAACAGATTACAGTTTAATGTT 2182
Qy 191 ---------- 191
Db 2183 GATAAATTAGCCATCGACCAACAACTAAAATATGTTGACAGCATTCAATTAGACACAGCT 2242
Qy 192 ---------AsnLeuAlaAspTrpArgTyrValVal---------- 200
Db 2243 CAAAGTAGCAATCTGAAATCCTATAGATATGTGTACACGAACAATAGCTCACTGGTTTTC 2302
Qy 201 ----------AspSerGluGlnTrpIleAsn 207
Db 2303 GACCCAAATGTAGCACCAGCAGAGGTTGACCTTAGTTCAGAATCTCTTAACTTGCTTAAT 2362
Qy 208 PheThrAsnAspThrThrTyrTyr---------- 215
Db 2363 TTTGATTCAGATGGCACCTATTTTTCTAATGCAAATAATAGACTTTTTTACACGCATTTA 2422
Qy 216 ----------ValArgIleArgValLeuArgProThrTyrAspVal--- 227
Db 2423 GGATATAGTGGCACACCAGGAGTTAACTATCTTCTCGTAATGTTTCTTTTTAACGCCAAA 2482
Qy 228 ProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAla 247
Db 2483 CCTGCGGATAAGTCAAAACTTGTC----------TACAAAGTCACTCGCAAACAA 2527
Qy 248 IleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGln 267
Db 2528 GTCACC----------GAAAACTTCGTGGAT----------GTCAACGGTGCC 2560
Qy 268 TyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHis 287
Db 2561 AAAATCACTGCACCAACAGGCTTCACCCAA---------GGTAACCAAGTACCAATGAAC 2611


Qy 288 ThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu---------- 300
Db 2612 AGTAACACCTTC---AAGTACACAGCGGCAAAAGCTTTACCAGCGACGTATACTACAGGT 2668
Qy 301 ------AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrPro 318
Db 2669 GGCAAAGTCTATACGTTCCAAGGGTGGTATAAAGGGAAAACCAAGCCAAGTACGTTG--- 2725
Qy 319 AlaTrpGluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhe 338
Db 2726 ----------AACAAAACAACAACT 2740
Qy 339 ProAlaGlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358
Db 2741 CCAACGTTCAATGCGACC----------TTTGATGGC 2767
Qy 359 AsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnVal 378
Db 2768 AATGACGATATGACCGCCATGTATAAGGAAGAAATACCAACAGCTAGTGTCACATTAACT 2827
Qy 379 AlaValAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPhe 398
Db 2828 CGACCAAAAGAAGTGATTGATACGAATACCAATGTAATCTGGACAACAACGATCACGAAT 2887
Qy 399 ThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeu 418
Db 2888 ACTAGCAAAGCACCC----------TTACAAAATCTCACCTTG 2920
Qy 419 GluThrArgProSer---SerArgArgLeuGluLeuPro------MetProProAlaAsp 435
Db 2921 AAAAAAGGGCCCAATTGGTCAGCTGGTCTGACGATCCCGACCTTTATGGAAGTGACACCA 2980
Qy 436 PheGlyGlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeu 455
Db 2981 GAAGGAGAAACGACAAAATCAATCCCA---GTAAATAGTACACTTTGGACAGAGGGGGTT 3037
Qy 456 GlyCysTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSer 475
Db 3038 ----------CCTTTACCAAATGCCGTTCCTATCGGC 3064
Qy 476 SerPheGlyAlaValSerPhe----------AsnAsnProGlyTyrGluArgThr 490
Db 3065 AAAAAAGTTTCAGTTGCTTTCACAACTCGCGCAACAGGGAAACCAAACACTGTTTTGAAA 3124
Qy 491 ArgAspLeuProAspTyrThrGlyIleArgAspSerPheAspGlnAsnMetSerThrAla 510
Db 3125 GCAGAAGTTGTAGTATTTGGTGGTATTAAAGATAGTACAGTGGATAACTTC--------- 3175
Qy 511 ValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLysThrTyrGlnGly 530
Db 3176 ---------GTGAGAATTCGTCCAAATGATCAAGAAGTAGTCACACCAACGACCGAAGGC 3226
Qy 531 TrpGluGlyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLeuLeu 550
Db 3227 TTCATCAGTGTGCCAACCTTCGAC------TTCGGCCAAGTGGGCGTTGCAGGAACTAAG 3280
Qy 551 LysAsnGluGluIleLeuCysLeuAlaAsp---------- 560
Db 3281 CAACAACACAGCTTGAAACAAGCCGCGGATTACTACGGTAACGGCACACGGAATCCGTAT 3340
Qy 561 ----------AspLeuAlaThrArgLeuThrGlyValTyr 570
Db 3341 CTGCGGATTAAGAAAACGCAACCCAATTGGAGCTTAACAGCGCAACTGTCACAACCAAAA 3400
Qy 571 ProAlaThrAspAsnPhe----------AlaAlaAlaValSer 581
Db 3401 TCAGCGACAGACAGCTTGCCTACAGCGACCCGCTTATTATTAGGGGCGGCGCCTGTCTCT 3460
Qy 582 AlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGlu------AlaThrSerSer 599
Db 3461 AGCTTTACCAATTACAATCAACCAACCGAGTTGAAAAATACGGTCGGTACCACGAGTGCC 3520
Qy 600 IleIleLysSerValGlyGluThrAla 608

```
Db   3521 ATTAGCTTAACAGCCAACAACACAGCA 3547

RESULT 6
ABN98194
ID   ABN98194 standard; DNA; 3695 BP.
XX
AC   ABN98194;
XX
DT   05-AUG-2002  (first entry)
XX
DE   E faecalis EF108 gene fragment.
XX
KW   Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
KW   gene; ds.
XX
OS   Enterococcus faecalis.
XX
PN   US2002045737-A1.
XX
PD   18-APR-2002.
XX
PF   04-MAY-1998;   98US-0071035.
XX
PR   04-MAY-1998;   98US-0071035.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Choi GH,  Bailey C,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 2002-425450/45.
DR   P-PSDB; ABP43438.
XX
PT   New genes and polypeptides from Enterococcus faecalis, useful as
PT   vaccines for preventing, treating or attenuating an infection caused by
PT   a member of the Enterococcus genus in an animal, particularly E.
PT   faecalis  -
XX
PS   Claim 1; Page 194-196; 255pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of polypeptides from Enterococcus faecalis. The proteins can be
CC   used as vaccines for preventing or attenuating an infection caused by a
CC   member of the Enterococcus genus in an animal, particularly E. faecalis.
CC   The polynucleotide is also useful for preventing or treating E. faecalis
CC   infection. The present sequence is a coding sequence of the invention.
XX
SQ   Sequence 3695 BP; 1293 A; 790 C; 739 G; 873 T; 0 other;

Alignment Scores:
Pred. No.:              0.131          Length:          3695
Score:                  139.50         Matches:         140
Percent Similarity:     31.38%         Conservative:    95
Best Local Similarity:  18.69%         Mismatches:      247
Query Match:            4.13%          Indels:          267
DB:                     24             Gaps:            31

US-09-677-653A-50 (1-647) x ABN98194 (1-3695)

Qy     19 AsnValArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArg 38
Db   1625 AACCGCAAAGTCACCGAGAACTTCGTAGACACCAACGGCGCTAAA-------------- 1669

Qy     39 ThrGlyArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGln 58
Db   1670 ------------ATCACACCGCCAACAGGTTTCACCCAAGGTAAAAAAACGGTGATTACA 1717

Qy     59 SerLeuAspAlaAsnThrVal------------ThrPheProAlaAsnIleSerSer--- 73
Db   1718 AGC---GACGCCTACACTTTCAAACAAGCAGGCACCTTACCAGACACTTACACAACAGGC 1774

Qy     74 ------MetProGluPheArgAsnTrpAlaLysGlyLysIleAspLeuAspSerAspSer 91
```

```
Db   1775 GGTAAGACCTACAAGTTCAAAGGTTGGTACAAAGGCAAGTCCATACTCAACACATTGACA 1834

Qy     92 IleGlyTrpTyrPheLysTyrLeuAspProAlaGlyAlaThrGluSerAlaArgAlaVal 111

Db   1835 ACT---------------------------------------------------------- 1837

Qy    112 GlyGluTyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArg 131
Db   1838 ---------ACCAAAGCGCCAAGTTATCAAGTGACCTACGATGACAATGATGATTTGAAT 1888

Qy    132 GluIleTyrAsnGluGluCysProValValThrAspValSerValProLeuAsp------ 149
Db   1889 GTGGTGTATGAAGAAGAAACAGTTACGACAGTGTATCCATCAGTCGATATGAACTTTGTG 1948

Qy    150 ----------------------------------------GlyArgGlnTrpSerLeu 155
Db   1949 AATGAAAAAGGCGGGGCTTTCACACCGGCGTTAACTTTTAGTGGTAAGTACTATGCGCAA 2008

Qy    156 SerIlePheSerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsn 175
Db   2009 AGTACGAGTGCGTAC------TTAAGAACCGATTTATATGACGTGACCTCAAAAAATAAT 2062

Qy    176 --------------------------LysGluMetSerLeuAspValValAsnAspLeu 186
Db   2063 GGTAATGGGCAATATACGGTAAGTATTAATAATGGTAGTATGCCATTGTCCCAAGAATTA 2122

Qy    187 IleGluTrpLeuAsn---------------------------------------------- 191
Db   2123 TTGAAAAAATATAATAATGGACAACCAATCAGTGCTACCAACAGATTACAGTTTAATGTT 2182

Qy    191 ---------------------------------------------------------- 191

Db   2183 GATAAATTAGCCATCGACCAACAACTAAAATATGTTGACAGCATTCAATTAGACACAGCT 2242

Qy    192 ---------AsnLeuAlaAspTrpArgTyrValVal----------------------- 200
Db   2243 CAAAGTAGCAATCTGAAATCCTATAGATATGTGTACACGAACAATAGCTCACTGGTTTTC 2302

Qy    201 ----------------------------------------AspSerGluGlnTrpIleAsn 207
Db   2303 GACCCAAATGTAGCACCAGCAGAGGTTGACCTTAGTTCAGAATCTCTTAACTTGCTTAAT 2362

Qy    208 PheThrAsnAspThrThrTyrTyr---------------------------------- 215
Db   2363 TTTGATTCAGATGGCACCTATTTTTCTAATGCAAATAATAGACTTTTTTACACGCATTTA 2422

Qy    216 ---------------------ValArgIleArgValLeuArgProThrTyrAspVal--- 227
Db   2423 GGATATAGTGGCACACCAGGAGTTAACTATCTTCTCGTAATGTTTCTTTTTAACGCCAAA 2482

Qy    228 ProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAla 247
Db   2483 CCTGCGGATAAGTCAAAACTTGTC--------------TACAAAGTCACTCGCAAACAA 2527

Qy    248 IleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGln 267
Db   2528 GTCACC--------------GAAAACTTCGTGGAT-----------GTCAACGGTGCC 2560

Qy    268 TyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHis 287
Db   2561 AAAATCACTGCACCAACAGGCTTCACCCAA---------GGTAACCAAGTACCAATGAAC 2611

Qy    288 ThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu--------------------- 300
Db   2612 AGTAACACCTTC---AAGTACACAGCGGCAAAAGCTTTACCAGCGACGTATACTACAGGT 2668

Qy    301 ------AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrPro 318
Db   2669 GGCAAAGTCTATACGTTCCAAGGGTGGTATAAAGGGAAAACCAAGCCAAGTACGTTG--- 2725

Qy    319 AlaTrpGluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhe 338
Db   2726 ----------------------------------------AACAAAACAACAACT 2740
```

```
Qy     339 ProAlaGlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358
Db    2741 CCAACGTTCAATGCGACC------------------------------TTTGATGGC 2767

Qy     359 AsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnVal 378
Db    2768 AATGACGATATGACCGCCATGTATAAGGAAGAAATACCAACAGCTAGTGTCACATTAACT 2827

Qy     379 AlaValAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPhe 398
Db    2828 CGACCAAAAGAAGTGATTGATACGAATACCAATGTAATCTGGACAACAACGATCACGAAT 2887

Qy     399 ThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeu 418
Db    2888 ACTAGCAAAGCACCC-------------------------TTACAAAATCTCACCTTG 2920

Qy     419 GluThrArgProSer---SerArgArgLeuGluLeuPro------MetProProAlaAsp 435
Db    2921 AAAAAAGGGCCCAATTGGTCAGCTGGTCTGACGATCCCGACCTTTATGGAAGTGACACCA 2980

Qy     436 PheGlyGlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeu 455
Db    2981 GAAGGAGAAACGACAAAATCAATCCCA---GTAAATAGTACACTTTGGACAGAGGGGGTT 3037

Qy     456 GlyCysTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSer 475
Db    3038 --------------------------------CCTTTACCAAATGCCGTTCCTATCGGC 3064

Qy     476 SerPheGlyAlaValSerPhe--------------AsnAsnProGlyTyrGluArgThr 490
Db    3065 AAAAAAGTTTCAGTTGCTTTCACAACTCGCGCAACAGGGAAACCAAACACTGTTTTGAAA 3124

Qy     491 ArgAspLeuProAspTyrThrGlyIleArgAspSerPheAspGlnAsnMetSerThrAla 510
Db    3125 GCAGAAGTTGTAGTATTTGGTGGTATTAAAGATAGTACAGTGGATAACTTC--------- 3175

Qy     511 ValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLysThrTyrGlnGly 530
Db    3176 ---------GTGAGAATTCGTCCAAATGATCAAGAAGTAGTCACACCAACGACCGAAGGC 3226

Qy     531 TrpGluGlyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLeuLeu 550
Db    3227 TTCATCAGTGTGCCAACCTTCGAC------TTCGGCCAAGTGGGCGTTGCAGGAACTAAG 3280

Qy     551 LysAsnGluGluIleLeuCysLeuAlaAsp---------------------------- 560
Db    3281 CAACAACACAGCTTGAAACAAGCCGCGGATTACTACGGTAACGGCACACGGAATCCGTAT 3340

Qy     561 ----------------------------AspLeuAlaThrArgLeuThrGlyValTyr 570
Db    3341 CTGCGGATTAAGAAAACGCAACCCAATTGGAGCTTAACAGCGCAACTGTCACAACCAAAA 3400

Qy     571 ProAlaThrAspAsnPhe-------------------------AlaAlaAlaValSer 581
Db    3401 TCAGCGACAGACAGCTTGCCTACAGCGACCCGCTTATTATTAGGGGCGGCGCCTGTCTCT 3460

Qy     582 AlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGlu------AlaThrSerSer 599
Db    3461 AGCTTTACCAATTACAATCAACCAACCGAGTTGAAAAATACGGTCGGTACCACGAGTGCC 3520

Qy     600 IleIleLysSerValGlyGluThrAla 608
Db    3521 ATTAGCTTAACAGCCAACAACACAGCA 3547

RESULT 7
AAX20208
ID   AAX20208 standard; DNA; 3840 BP.
XX
AC   AAX20208;
XX
DT   20-APR-1999  (first entry)
XX
DE   Enterococcus faecalis gene EF108.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic; ss.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C,  Choi GH,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   P-PSDB; AAY00218.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 1; Page 208-209; 301pp; English.
XX
CC   The present sequence represents a gene isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
XX
SQ   Sequence 3840 BP; 1335 A; 815 C; 778 G; 912 T; 0 other;

Alignment Scores:
Pred. No.:                0.138          Length:          3840
Score:                    139.50         Matches:         140
Percent Similarity:       31.38%         Conservative:    95
Best Local Similarity:    18.69%         Mismatches:      247
Query Match:              4.13%          Indels:          267
DB:                       20             Gaps:            31

US-09-677-653A-50 (1-647) x AAX20208 (1-3840)

Qy      19 AsnValArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArg 38
Db    1762 AACCGCAAAGTCACCGAGAACTTCGTAGACACCAACGGCGCTAAA-------------- 1806

Qy      39 ThrGlyArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGln 58
Db    1807 ------------ATCACACCGCCAACAGGTTTCACCCAAGGTAAAAAAACGGTGATTACA 1854

Qy      59 SerLeuAspAlaAsnThrVal------------ThrPheProAlaAsnIleSerSer--- 73
Db    1855 AGC---GACGCCTACACTTTCAAACAAGCAGGCACCTTACCAGACACTTACACAACAGGC 1911

Qy      74 ------MetProGluPheArgAsnTrpAlaLysGlyLysIleAspLeuAspSerAspSer 91
Db    1912 GGTAAGACCTACAAGTTCAAAGGTTGGTACAAAGGCAAGTCCATACTCAACACATTGACA 1971

Qy      92 IleGlyTrpTyrPheLysTyrLeuAspProAlaGlyAlaThrGluSerAlaArgAlaVal 111
Db    1972 ACT---------------------------------------------------------- 1974
```

```
Qy    112 GlyGluTyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArg 131
Db   1975 ---------ACCAAAGCGCCAAGTTATCAAGTGACCTACGATGACAATGATGATTTGAAT 2025

Qy    132 GluIleTyrAsnGluGluCysProValValThrAspValSerValProLeuAsp------ 149
Db   2026 GTGGTGTATGAAGAAGAAACAGTTACGACAGTGTATCCATCAGTCGATATGAACTTTGTG 2085

Qy    150 ----------------------------------------GlyArgGlnTrpSerLeu 155
Db   2086 AATGAAAAAGGCGGGGCTTTCACACCGGCGTTAACTTTTAGTGGTAAGTACTATGCGCAA 2145

Qy    156 SerIlePheSerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsn 175
Db   2146 AGTACGAGTGCGTAC------TTAAGAACCGATTTATATGACGTGACCTCAAAAAATAAT 2199

Qy    176 --------------------------LysGluMetSerLeuAspValValAsnAspLeu 186
Db   2200 GGTAATGGGCAATATACGGTAAGTATTAATAATGGTAGTATGCCATTGTCCCAAGAATTA 2259

Qy    187 IleGluTrpLeuAsn---------------------------------------------- 191
Db   2260 TTGAAAAAATATAATAATGGACAACCAATCAGTGCTACCAACAGATTACAGTTTAATGTT 2319

Qy    191 ---------------------------------------------------------- 191
Db   2320 GATAAATTAGCCATCGACCAACAACTAAAATATGTTGACAGCATTCAATTAGACACAGCT 2379

Qy    192 ---------AsnLeuAlaAspTrpArgTyrValVal----------------------- 200
Db   2380 CAAAGTAGCAATCTGAAATCCTATAGATATGTGTACACGAACAATAGCTCACTGGTTTTC 2439

Qy    201 ---------------------------------------AspSerGluGlnTrpIleAsn 207
Db   2440 GACCCAAATGTAGCACCAGCAGAGGTTGACCTTAGTTCAGAATCTCTTAACTTGCTTAAT 2499

Qy    208 PheThrAsnAspThrThrTyrTyr----------------------------------- 215
Db   2500 TTTGATTCAGATGGCACCTATTTTTCTAATGCAAATAATAGACTTTTTTACACGCATTTA 2559

Qy    216 ---------------------ValArgIleArgValLeuArgProThrTyrAspVal--- 227
Db   2560 GGATATAGTGGCACACCAGGAGTTAACTATCTTCTCGTAATGTTTCTTTTTAACGCCAAA 2619

Qy    228 ProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAla 247
Db   2620 CCTGCGGATAAGTCAAAACTTGTC---------------TACAAAGTCACTCGCAAACAA 2664

Qy    248 IleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGln 267
Db   2665 GTCACC---------------GAAAACTTCGTGGAT------------GTCAACGGTGCC 2697

Qy    268 TyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHis 287
Db   2698 AAAATCACTGCACCAACAGGCTTCACCCAA---------GGTAACCAAGTACCAATGAAC 2748

Qy    288 ThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu---------------------- 300
Db   2749 AGTAACACCTTC---AAGTACACAGCGGCAAAAGCTTTACCAGCGACGTATACTACAGGT 2805

Qy    301 ------AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrPro 318
Db   2806 GGCAAAGTCTATACGTTCCAAGGGTGGTATAAAGGGAAAACCAAGCCAAGTACGTTG--- 2862

Qy    319 AlaTrpGluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhe 338
Db   2863 ----------------------------------------------AACAAAACAACAACT 2877

Qy    339 ProAlaGlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358
Db   2878 CCAACGTTCAATGCGACC------------------------------TTTGATGGC 2904

Qy    359 AsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnVal 378

Db   2905 AATGACGATATGACCGCCATGTATAAGGAAGAAATACCAACAGCTAGTGTCACATTAACT 2964

Qy    379 AlaValAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPhe 398
Db   2965 CGACCAAAAGAAGTGATTGATACGAATACCAATGTAATCTGGACAACAACGATCACGAAT 3024

Qy    399 ThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeu 418
Db   3025 ACTAGCAAAGCACCC--------------------------TTACAAAATCTCACCTTG 3057

Qy    419 GluThrArgProSer---SerArgArgLeuGluLeuPro------MetProProAlaAsp 435
Db   3058 AAAAAAGGGCCCAATTGGTCAGCTGGTCTGACGATCCCGACCTTTATGGAAGTGACACCA 3117

Qy    436 PheGlyGlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeu 455
Db   3118 GAAGGAGAAACGACAAAATCAATCCCA---GTAAATAGTACACTTTGGACAGAGGGGGTT 3174

Qy    456 GlyCysTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSer 475
Db   3175 ------------------------------CCTTTACCAAATGCCGTTCCTATCGGC 3201

Qy    476 SerPheGlyAlaValSerPhe--------------AsnAsnProGlyTyrGluArgThr 490
Db   3202 AAAAAAGTTTCAGTTGCTTTCACAACTCGCGCAACAGGGAAACCAAACACTGTTTTGAAA 3261

Qy    491 ArgAspLeuProAspTyrThrGlyIleArgAspSerPheAspGlnAsnMetSerThrAla 510
Db   3262 GCAGAAGTTGTAGTATTTGGTGGTATTAAAGATAGTACAGTGGATAACTTC--------- 3312

Qy    511 ValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLysThrTyrGlnGly 530
Db   3313 ---------GTGAGAATTCGTCCAAATGATCAAGAAGTAGTCACACCAACGACCGAAGGC 3363

Qy    531 TrpGluGlyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLeuLeu 550
Db   3364 TTCATCAGTGTGCCAACCTTCGAC------TTCGGCCAAGTGGGCGTTGCAGGAACTAAG 3417

Qy    551 LysAsnGluGluIleLeuCysLeuAlaAsp------------------------------ 560
Db   3418 CAACAACACAGCTTGAAACAAGCCGCGGATTACTACGGTAACGGCACACGGAATCCGTAT 3477

Qy    561 -----------------------------AspLeuAlaThrArgLeuThrGlyValTyr 570
Db   3478 CTGCGGATTAAGAAAACGCAACCCAATTGGAGCTTAACAGCGCAACTGTCACAACCAAAA 3537

Qy    571 ProAlaThrAspAsnPhe--------------------------AlaAlaAlaValSer 581
Db   3538 TCAGCGACAGACAGCTTGCCTACAGCGACCCGCTTATTATTAGGGGCGGCGCCTGTCTCT 3597

Qy    582 AlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGlu------AlaThrSerSer 599
Db   3598 AGCTTTACCAATTACAATCAACCAACCGAGTTGAAAAATACGGTCGGTACCACGAGTGCC 3657

Qy    600 IleIleLysSerValGlyGluThrAla 608
Db   3658 ATTAGCTTAACAGCCAACAACACAGCA 3684

RESULT 8
ABN98193
ID   ABN98193 standard; DNA; 3840 BP.
XX
AC   ABN98193;
XX
DT   05-AUG-2002  (first entry)
XX
DE   E faecalis EF108 gene.
XX
KW   Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
KW   gene; ds.
XX
OS   Enterococcus faecalis.
```

XX
PN US2002045737-A1.
XX
PD 18-APR-2002.
XX
PF 04-MAY-1998; 98US-0071035.
XX
PR 04-MAY-1998; 98US-0071035.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR WPI; 2002-425450/45.
DR P-PSDB; ABP43437.
XX
PT New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis -
XX
PS Claim 1; Page 192-193; 255pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of polypeptides from Enterococcus faecalis. The proteins can be
CC used as vaccines for preventing or attenuating an infection caused by a
CC member of the Enterococcus genus in an animal, particularly E. faecalis.
CC The polynucleotide is also useful for preventing or treating E. faecalis
CC infection. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 3840 BP; 1335 A; 815 C; 778 G; 912 T; 0 other;

Alignment Scores:

Pred. No.:	0.138	Length:	3840
Score:	139.50	Matches:	140
Percent Similarity:	31.38%	Conservative:	95
Best Local Similarity:	18.69%	Mismatches:	247
Query Match:	4.13%	Indels:	267
DB:	24	Gaps:	31

US-09-677-653A-50 (1-647) x ABN98193 (1-3840)

```
Qy      19 AsnValArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArg 38
           |||   :::|||:::   |||   |||      ||||||   :::
Db    1762 AACCGCAAAGTCACCGAGAACTTCGTAGACACCAACGGCGCTAAA---------------- 1806

Qy      39 ThrGlyArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGln 58
                   ::::::||||||      ||||||         :::   :::
Db    1807 ------------ATCACACCGCCAACAGGTTTCACCCAAGGTAAAAAAACGGTGATTACA 1854

Qy      59 SerLeuAspAlaAsnThrVal------------ThrPheProAlaAsnIleSerSer--- 73
           |||   ||||||   |||               |||   |||        ::::::
Db    1855 AGC---GACGCCTACACTTTCAAACAAGCAGGCACCTTACCAGACACTTACACAACAGGC 1911

Qy      74 ------MetProGluPheArgAsnTrpAlaLysGlyLysIleAspLeuAspSerAspSer 91
                 :::|||:::   |||   |||||||||      |||::::::   :::
Db    1912 GGTAAGACCTACAAGTTCAAAGGTTGGTACAAAGGCAAGTCCATACTCAACACATTGACA 1971

Qy      92 IleGlyTrpTyrPheLysTyrLeuAspProAlaGlyAlaThrGluSerAlaArgAlaVal 111

Db    1972 ACT---------------------------------------------------------- 1974

Qy     112 GlyGluTyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArg 131
                ::::|||   |||         |||   :::      :::   ::::::
Db    1975 ---------ACCAAAGCGCCAAGTTATCAAGTGACCTACGATGACAATGATGATTTGAAT 2025

Qy     132 GluIleTyrAsnGluGluCysProValValThrAspValSerValProLeuAsp------ 149
              :::|||   ||||||                ||||||   ::::::
Db    2026 GTGGTGTATGAAGAAGAAACAGTTACGACAGTGTATCCATCAGTCGATATGAACTTTGTG 2085

Qy     150 ----------------------------------------GlyArgGlnTrpSerLeu 155
                                                     |||:::   ::::::
Db    2086 AATGAAAAAGGCGGGGCTTTCACACCGGCGTTAACTTTTAGTGGTAAGTACTATGCGCAA 2145

Qy     156 SerIlePheSerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsn 175
           |||      ::::::         ||||||            |||   :::      |||
Db    2146 AGTACGAGTGCGTAC------TTAAGAACCGATTTATATGACGTGACCTCAAAAAATAAT 2199

Qy     176 --------------------------LysGluMetSerLeuAspValValAsnAspLeu 186
                                       |||:::  :::      :::|||
Db    2200 GGTAATGGGCAATATACGGTAAGTATTAATAATGGTAGTATGCCATTGTCCCAAGAATTA 2259

Qy     187 IleGluTrpLeuAsn---------------------------------------------- 191
           ::::::      |||
Db    2260 TTGAAAAAATATAATAATGGACAACCAATCAGTGCTACCAACAGATTACAGTTTAATGTT 2319

Qy     191 ------------------------------------------------------------ 191

Db    2320 GATAAATTAGCCATCGACCAACAACTAAAATATGTTGACAGCATTCAATTAGACACAGCT 2379

Qy     192 ---------AsnLeuAlaAspTrpArgTyrValVal------------------------ 200
                    ||||||      :::||||||||||
Db    2380 CAAAGTAGCAATCTGAAATCCTATAGATATGTGTACACGAACAATAGCTCACTGGTTTTC 2439

Qy     201 ---------------------------------------AspSerGluGlnTrpIleAsn 207
                                                    :::|||         :::|||
Db    2440 GACCCAAATGTAGCACCAGCAGAGGTTGACCTTAGTTCAGAATCTCTTAACTTGCTTAAT 2499

Qy     208 PheThrAsnAspThrThrTyrTyr---------------------------------- 215
           |||   :::|||   ||||||:::
Db    2500 TTTGATTCAGATGGCACCTATTTTTCTAATGCAAATAATAGACTTTTTTACACGCATTTA 2559

Qy     216 --------------------ValArgIleArgValLeuArgProThrTyrAspVal--- 227
                               |||         ::::::      ::::::
Db    2560 GGATATAGTGGCACACCAGGAGTTAACTATCTTCTCGTAATGTTTCTTTTTAACGCCAAA 2619

Qy     228 ProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAla 247
           |||               ||||||               |||:::::|||   |||
Db    2620 CCTGCGGATAAGTCAAAACTTGTC--------------TACAAAGTCACTCGCAAACAA 2664

Qy     248 IleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGln 267
           :::|||                        ||||||      :::   |||
Db    2665 GTCACC--------------GAAAACTTCGTGGAT-----------GTCAACGGTGCC 2697

Qy     268 TyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHis 287
                   ||||||         |||                        ::::::
Db    2698 AAAATCACTGCACCAACAGGCTTCACCCAA---------GGTAACCAAGTACCAATGAAC 2748

Qy     288 ThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu---------------------- 300
           :::   ||||||   :::   :::::|||   ||||||
Db    2749 AGTAACACCTTC---AAGTACACAGCGGCAAAAGCTTTACCAGCGACGTATACTACAGGT 2805

Qy     301 ------AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrPro 318
                 :::   :::   |||       :::|||   |||   |||:::
Db    2806 GGCAAAGTCTATACGTTCCAAGGGTGGTATAAAGGGAAAACCAAGCCAAGTACGTTG--- 2862

Qy     319 AlaTrpGluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhe 338
                                                     |||   ||||||
Db    2863 ---------------------------------------------AACAAAACAACAACT 2877

Qy     339 ProAlaGlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358
           |||      :::   :::                                  :::|||
Db    2878 CCAACGTTCAATGCGACC-------------------------------TTTGATGGC 2904

Qy     359 AsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnVal 378
           ||||||               :::               |||   :::
Db    2905 AATGACGATATGACCGCCATGTATAAGGAAGAAATACCAACAGCTAGTGTCACATTAACT 2964

Qy     379 AlaValAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPhe 398
                 |||||||:::   ||||||                  :::      |||
Db    2965 CGACCAAAAGAAGTGATTGATACGAATACCAATGTAATCTGGACAACAACGATCACGAAT 3024

Qy     399 ThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeu 418
           |||   :::   |||                              :::|||      |||
Db    3025 ACTAGCAAAGCACCC--------------------------TTACAAAATCTCACCTTG 3057
```

```
Qy    419 GluThrArgProSer---SerArgArgLeuGluLeuPro------MetProProAlaAsp 435
Db   3058 AAAAAAGGGCCCAATTGGTCAGCTGGTCTGACGATCCCGACCTTTATGGAAGTGACACCA 3117

Qy    436 PheGlyGlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeu 455
Db   3118 GAAGGAGAAACGACAAAATCAATCCCA---GTAAATAGTACACTTTGGACAGAGGGGGTT 3174

Qy    456 GlyCysTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSer 475
Db   3175 ------------------------------CCTTTACCAAATGCCGTTCCTATCGGC 3201

Qy    476 SerPheGlyAlaValSerPhe--------------AsnAsnProGlyTyrGluArgThr 490
Db   3202 AAAAAAGTTTCAGTTGCTTTCACAACTCGCGCAACAGGGAAACCAAACACTGTTTTGAAA 3261

Qy    491 ArgAspLeuProAspTyrThrGlyIleArgAspSerPheAspGlnAsnMetSerThrAla 510
Db   3262 GCAGAAGTTGTAGTATTTGGTGGTATTAAAGATAGTACAGTGGATAACTTC--------- 3312

Qy    511 ValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLysThrTyrGlnGly 530
Db   3313 ---------GTGAGAATTCGTCCAAATGATCAAGAAGTAGTCACACCAACGACCGAAGGC 3363

Qy    531 TrpGluGlyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLeuLeu 550
Db   3364 TTCATCAGTGTGCCAACCTTCGAC------TTCGGCCAAGTGGGCGTTGCAGGAACTAAG 3417

Qy    551 LysAsnGluGluIleLeuCysLeuAlaAsp---------------------------- 560
Db   3418 CAACAACACAGCTTGAAACAAGCCGCGGATTACTACGGTAACGGCACACGGAATCCGTAT 3477

Qy    561 -----------------------------AspLeuAlaThrArgLeuThrGlyValTyr 570
Db   3478 CTGCGGATTAAGAAAACGCAACCCAATTGGAGCTTAACAGCGCAACTGTCACAACCAAAA 3537

Qy    571 ProAlaThrAspAsnPhe--------------------------AlaAlaAlaValSer 581
Db   3538 TCAGCGACAGACAGCTTGCCTACAGCGACCCGCTTATTATTAGGGGCGGCGCCTGTCTCT 3597

Qy    582 AlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGlu------AlaThrSerSer 599
Db   3598 AGCTTTACCAATTACAATCAACCAACCGAGTTGAAAAATACGGTCGGTACCACGAGTGCC 3657

Qy    600 IleIleLysSerValGlyGluThrAla 608
Db   3658 ATTAGCTTAACAGCCAACAACACAGCA 3684
```

RESULT 9
AAX13104
ID AAX13104 standard; DNA; 19031 BP.
XX
AC AAX13104;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:167.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN WO9850555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 939-949; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 19031 BP; 6046 A; 3825 C; 4068 G; 5072 T; 20 other;

Alignment Scores:

Pred. No.:	1.26	Length:	19031
Score:	139.50	Matches:	140
Percent Similarity:	31.38%	Conservative:	95
Best Local Similarity:	18.69%	Mismatches:	247
Query Match:	4.13%	Indels:	267
DB:	20	Gaps:	31

US-09-677-653A-50 (1-647) x AAX13104 (1-19031)

```
Qy     19 AsnValArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArg 38
Db   4462 AACCGCAAAGTCACCGAGAACTTCGTAGACACCAACGGCGCTAAA-------------- 4506

Qy     39 ThrGlyArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGln 58
Db   4507 ------------ATCACACCGCCAACAGGTTTCACCCAAGGTAAAAAAACGGTGATTACA 4554

Qy     59 SerLeuAspAlaAsnThrVal------------ThrPheProAlaAsnIleSerSer--- 73
Db   4555 AGC---GACGCCTACACTTTCAAACAAGCAGGCACCTTACCAGACACTTACACAACAGGC 4611

Qy     74 ------MetProGluPheArgAsnTrpAlaLysGlyLysIleAspLeuAspSerAspSer 91
Db   4612 GGTAAGACCTACAAGTTCAAAGGTTGGTACAAAGGCAAGTCCATACTCAACACATTGACA 4671

Qy     92 IleGlyTrpTyrPheLysTyrLeuAspProAlaGlyAlaThrGluSerAlaArgAlaVal 111
Db   4672 ACT---------------------------------------------------------- 4674

Qy    112 GlyGluTyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArg 131
Db   4675 ---------ACCAAAGCGCCAAGTTATCAAGTGACCTACGATGACAATGATGATTTGAAT 4725

Qy    132 GluIleTyrAsnGluGluCysProValValThrAspValSerValProLeuAsp------ 149
Db   4726 GTGGTGTATGAAGAAGAAACAGTTACGACAGTGTATCCATCAGTCGATATGAACTTTGTG 4785

Qy    150 ----------------------------------------GlyArgGlnTrpSerLeu 155
Db   4786 AATGAAAAAGGCGGGGCTTTCACACCGGCGTTAACTTTTAGTGGTAAGTACTATGCGCAA 4845

Qy    156 SerIlePheSerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsn 175
```

```
Db   4846 AGTACGAGTGCGTAC------TTAAGAACCGATTTATATGACGTGACCTCAAAAAATAAT 4899

Qy    176 -------------------------LysGluMetSerLeuAspValValAsnAspLeu 186

Db   4900 GGTAATGGGCAATATACGGTAAGTATTAATAATGGTAGTATGCCATTGTCCCAAGAATTA 4959

Qy    187 IleGluTrpLeuAsn---------------------------------------------- 191

Db   4960 TTGAAAAAATATAATAATGGACAACCAATCAGTGCTACCAACAGATTACAGTTTAATGTT 5019

Qy    191 ------------------------------------------------------------ 191

Db   5020 GATAAATTAGCCATCGACCAACAACTAAAATATGTTGACAGCATTCAATTAGACACAGCT 5079

Qy    192 ---------AsnLeuAlaAspTrpArgTyrValVal----------------------- 200

Db   5080 CAAAGTAGCAATCTGAAATCCTATAGATATGTGTACACGAACAATAGCTCACTGGTTTTC 5139

y     201 ------------------------------------AspSerGluGlnTrpIleAsn 207

Db   5140 GACCCAAATGTAGCACCAGCAGAGGTTGACCTTAGTTCAGAATCTCTTAACTTGCTTAAT 5199

Qy    208 PheThrAsnAspThrThrTyrTyr---------------------------------- 215

Db   5200 TTTGATTCAGATGGCACCTATTTTTCTAATGCAAATAATAGACTTTTTTACACGCATTTA 5259

Qy    216 --------------------ValArgIleArgValLeuArgProThrTyrAspVal--- 227

Db   5260 GGATATAGTGGCACACCAGGAGTTAACTATCTTCTCGTAATGTTTCTTTTTAACGCCAAA 5319

Qy    228 ProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAla 247

Db   5320 CCTGCGGATAAGTCAAAACTTGTC--------------TACAAAGTCACTCGCAAACAA 5364

Qy    248 IleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGln 267

Db   5365 GTCACC--------------GAAAACTTCGTGGAT-----------GTCAACGGTGCC 5397

Qy    268 TyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHis 287

Db   5398 AAAATCACTGCACCAACAGGCTTCACCCAA---------GGTAACCAAGTACCAATGAAC 5448

Qy    288 ThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu--------------------- 300

Db   5449 AGTAACACCTTC---AAGTACACAGCGGCAAAAGCTTTACCAGCGACGTATACTACAGGT 5505

Qy    301 ------AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrPro 318

Db   5506 GGCAAAGTCTATACGTTCCAAGGGTGGTATAAAGGGAAAACCAAGCCAAGTACGTTG--- 5562

Qy    319 AlaTrpGluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhe 338

Db   5563 ---------------------------------------------AACAAAACAACAACT 5577

Qy    339 ProAlaGlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358

Db   5578 CCAACGTTCAATGCGACC-------------------------------TTTGATGGC 5604

Qy    359 AsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnVal 378

Db   5605 AATGACGATATGACCGCCATGTATAAGGAAGAAATACCAACAGCTAGTGTCACATTAACT 5664

Qy    379 AlaValAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPhe 398

Db   5665 CGACCAAAAGAAGTGATTGATACGAATACCAATGTAATCTGGACAACAACGATCACGAAT 5724

Qy    399 ThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeu 418

Db   5725 ACTAGCAAAGCACCC-------------------------TTACAAAATCTCACCTTG 5757

Qy    419 GluThrArgProSer---SerArgArgLeuGluLeuPro------MetProProAlaAsp 435

Db   5758 AAAAAAGGGCCCAATTGGTCAGCTGGTCTGACGATCCCGACCTTTATGGAAGTGACACCA 5817

Qy    436 PheGlyGlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeu 455

Db   5818 GAAGGAGAAACGACAAAATCAATCCCA---GTAAATAGTACACTTTGGACAGAGGGGGTT 5874

Qy    456 GlyCysTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSer 475

Db   5875 -------------------------------CCTTTACCAAATGCCGTTCCTATCGGC 5901

Qy    476 SerPheGlyAlaValSerPhe--------------AsnAsnProGlyTyrGluArgThr 490

Db   5902 AAAAAAGTTTCAGTTGCTTTCACAACTCGCGCAACAGGGAAACCAAACACTGTTTTGAAA 5961

Qy    491 ArgAspLeuProAspTyrThrGlyIleArgAspSerPheAspGlnAsnMetSerThrAla 510

Db   5962 GCAGAAGTTGTAGTATTTGGTGGTATTAAAGATAGTACAGTGGATAACTTC--------- 6012

Qy    511 ValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLysThrTyrGlnGly 530

Db   6013 --------GTGAGAATTCGTCCAAATGATCAAGAAGTAGTCACACCAACGACCGAAGGC 6063

Qy    531 TrpGluGlyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLeuLeu 550

Db   6064 TTCATCAGTGTGCCAACCTTCGAC------TTCGGCCAAGTGGGCGTTGCAGGAACTAAG 6117

Qy    551 LysAsnGluGluIleLeuCysLeuAlaAsp------------------------------ 560

Db   6118 CAACAACACAGCTTGAAACAAGCCGCGGATTACTACGGTAACGGCACACGGAATCCGTAT 6177

Qy    561 -----------------------------AspLeuAlaThrArgLeuThrGlyValTyr 570

Db   6178 CTGCGGATTAAGAAAACGCAACCCAATTGGAGCTTAACAGCGCAACTGTCACAACCAAAA 6237

Qy    571 ProAlaThrAspAsnPhe--------------------------AlaAlaAlaValSer 581

Db   6238 TCAGCGACAGACAGCTTGCCTACAGCGACCCGCTTATTATTAGGGGCGGCGCCTGTCTCT 6297

Qy    582 AlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGlu------AlaThrSerSer 599

Db   6298 AGCTTTACCAATTACAATCAACCAACCGAGTTGAAAAATACGGTCGGTACCACGAGTGCC 6357

Qy    600 IleIleLysSerValGlyGluThrAla 608

Db   6358 ATTAGCTTAACAGCCAACAACACAGCA 6384
```

RESULT 10
ABL30182
ID ABL30182 standard; DNA; 8496 BP.
XX
AC ABL30182;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 42019.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;

```
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 42019; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8496 BP; 2467 A; 1839 C; 1860 G; 2330 T; 0 other;

Alignment Scores:
Pred. No.:                2.37          Length:          8496
Score:                    130.50        Matches:         148
Percent Similarity:       31.13%        Conservative:    73
Best Local Similarity:    20.85%        Mismatches:      256
Query Match:              3.87%         Indels:          233
DB:                       23            Gaps:            36

US-09-677-653A-50 (1-647) x ABL30182 (1-8496)

Qy      70 AsnIleSerSerMetProGluPheArgAsnTrpAlaLysGlyLysIleAspLeuAspSer 89
           |||::::::      |||   |||              |||::::::
Db    4593 AACGTAACCATTAAGCCCACTTTCCTCACTACTCCCTATGGTCGGGTG------------ 4640

Qy      90 AspSerIleGlyTrpTyrPheLysTyrLeuAspProAlaGlyAla---------ThrGlu 106
                 ||||||   |||:::|||      |||             ||||||
Db    4641 --------------TACTTCTACTACGTGGATGAGACAGGAGAGTTCCGATATACCGAG 4685

Qy     107 SerAlaArgAlaValGlyGluTyrSerLysIleProAspGlyLeuValLysPheSerVal 126
                                                              |||||||||
Db    4686 GAAACA----------------------------------------------TTCAGTGTG 4700

Qy     127 AspAlaGluIleArgGluIleTyrAsnGluGluCysProValValThrAspValSerVal 146
           :::   |||::::::               :::   |||
Db    4701 GAGGTCGAGCTGCAGAACCAGATCGAAATCAAGGCTCCCGCGGAAGTTAAA--------- 4751

Qy     147 ProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPheProMetPheArgThrAla 166
                 |||         :::|||   |||   :::   |||           :::
Db    4752 ------CCGGGTGCCGACGTTGCACTGGAGATCAAGACGTCGCCA---------AAGTCA 4796

Qy     167 TyrValAlaValAlaAsnValGluAsnLysGluMetSerLeuAspValValAsnAspLeu 186
           :::|||   :::      |||:::         :::   |||         |||||||||
Db    4797 TTCGTTGGATTGTTGGCCGTCGATCAGAGTGTCCTTCTGCTGGGCAGCAACAACGATTTG 4856

Qy     187 ------------IleGluTrpLeuAsnAsnLeuAlaAspTrpArgTyrValValAspSer 202
                       |||      |||      |||                      |||
Db    4857 AACAAGGAATCCTTCAACTGGCGCCTAAATGGCTATGAC--------------ACCTCA 4901

Qy     203 GluGlnTrpIleAsnPheThrAsnAspThrThrTyrTyrValArgIleArgValLeuArg 222
                |||                :::|||||||
Db    4902 ACTCCCTGG---------CAGGGTGGATACTCCTACTAC-------------------- 4931

Qy     223 ProThrTyrAspValProAspProThrGluGlyLeuVal-------------------- 235
                       |||            |||:::|||
Db    4932 --------------CCCGGAGAGCGAACTGGAGTCGTGACTATGACCAATGCCTATTTC 4976

Qy     236 ---------ArgThrValSerAspTyrArgLeu---ThrTyrLys------AlaIleThr 249
                   ||||||      ||||||   :::   |||||||||          |||
Db    4977 TTCTATAATCGCACCGCCCCAGACTACAATATTCGTACGTATAAATTTGTCATAATAATA 5036

Qy     250 CysGluAlaAsnMetProThrLeu---ValAspGlnGlyPheTrpIleGlyGlyGlnTyr 268
                  |||:::      |||   |||   :::||||||      |||      :::
Db    5037 AACCACATTAATTTGCAATTATTGGAAGTTACAGAAGGATTTGGT---GGAAGCTCTTTC 5093

Qy     269 AlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHisThr 288
           |||:::      |||::::::                :::::::::      |||:::
Db    5094 GCCATGAGAAAGACTACAGTG------------GCACATGATAGTCATGTATTCCATTCC 5141

Qy     289 ------------------------LeuThrPheAlaArgProSerSerAlaAlaAlaLeu 300
                                    :::   |||:::      |||:::::::|||:::
Db    5142 GGTGCTGGCGGACCAACTCAGGCCGTGGGATTCTCCGCCGAAAGTGCGTCTGCATCTGCA 5201

Qy     301 Ala---------------------------------------------------------- 301
           |||
Db    5202 GCACCAGTAGTGCGAAAGAATTTCGCAGAGACCTGGATCTTTGCCGATATCGAGAGCACC 5261

Qy     302 ------------PheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThr 317
                       |||   |||                     |||   |||
Db    5262 GAGGAGGAGGTGTTCAAGTGGGTA--------------AAGACCATACCCGATACGATC 5306

Qy     318 ProAlaTrp---------------------GluGlnAlaSerSerGlyGlyTyrLeuThr 330
           |||                                    |||
Db    5307 ACCAACTGGGTGGTCACCGGCTTTTCACTGCATCCCCAGAAGGGATTAGGTGTCACCAAT 5366

Qy     331 TrpArgHisAsnGlyThrThrProAlaGlySerValSerTyrValLeuProGluGly 350
           :::   |||      ||||||            ||||||      ||||||
Db    5367 GATCAGACAAATATCAAGACATTCCAGCCGTTCTTTGTCTCCGTTCGATTGCCA------ 5420

Qy     351 PheAlaLeuGluArg-----------------------------------TyrAspPro 358
           ::::::::::::|||                                    |||   |||
Db    5421 TACTCCGTGAAGCGAGGTGAGGTGATCAATGTACCGGCACTGGTCTTTAACTACCTGCCC 5480

Qy     359 ----------------------------AsnAspGlySerTrpThrAspPheAlaSer 368
                                         |||:::          ||||||
Db    5481 AAGACACTGGACGTGGAACTGACCCTGGACAACGAGGATCAAGAGTATGACTTCGTGGAT 5540

Qy     369 AlaGlyAspThrValThrPheArgGlnValAlaValAspGluValValValThrAsnAsn 388
           |||   :::   |||         |||               :::   |||      |||
Db    5541 GCCTCGAATGAGGTTATTGGTGACCAGAAGCGTACGCAGAACATCAGAGTGGGCGCCAAC 5600

Qy     389 ProAlaGlyGlyGlySerAlaProThrPheThrValArg---------------------- 401
           |||   |||   |||         |||   :::|||
Db    5601 GAAGCGGCAGGAGCCTCA---------TTTTTGATCCGACCCAAGGTCATTGGAAATATT 5651

Qy     401 ---------------------------------------------------------- 401

Db    5652 TTGCTGAAATTCAAGGCCATCTCACCGCTGGCCGGAGATGCTATACACAAGCCGCTAAAG 5711

Qy     402 ---ValProProSer---AsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeuGlu 419
                |||||               |||   |||      |||         |||   :::
Db    5712 GTAGTGCCCGAAGGAATTACCCAGTACCAGAACAGGGCATTCTTTATCAATCTTAAGGAT 5771

Qy     420 ThrArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThr 439
           |||                  ||||||   :::|||
Db    5772 ACGGGTGAGTTTAAGAACACCTTCGAGCTGGAAGTGCCGGAG------------GATGTG 5819

Qy     440 ValAlaAsnAsnProLysIleGlu----------------------------------- 447
           |||   ::::::   ::::::|||
Db    5820 GTCCCCGACTCCGAGCGTGTGGAATTCGGATTGGTGGGTGATCTCTTGGGACCAGTGGTC 5879

Qy     448 ---------GlnSerLeuLeuLysGluThrLeuGlyCys---------------------- 457
                   ::::::||||||:::          ||||||
Db    5880 AAGAACCTTGAGAACCTGTTGCGATTGCCCAGCGGTTGCGGTGAGCAGACCATGTCCAAA 5939

Qy     458 ------------TyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrPro 473
                       |||||||||          :::::::::   :::   :::|||||||||
Db    5940 TTGGTGCCCAACTATCTGGTGAGGGATTATCTGAAGAGC---ATCAAGAAACTCACACCT 5996

Qy     474 AlaSerSerPheGlyAlaValSerPheAsnAsnProGlyTyrGluArgThrArgAspLeu 493
           |||                                  ||||||:::
Db    5997 GCATTGGACACTCGCATCAAGAGGAATCTGCAGGACGGATACCAACATATGTTGCACTAT 6056

Qy     494 ProAspTyrThrGlyIleArgAspSerPheAsp-------------------------- 504
```

```
                    |||         ||||||
Db   6057 CGCCACGACGACGGCAGCTTCAGTTCCTTTGGGCCCACCAAATGGAGGCAGGAGGATCCA 6116

Qy    505 ---GlnAsnMetSerThrAlaValAlaHisPhe------ArgSerLeuSerHisSerCys 521
          :::|||   ||||||   :::      :::      ||||||   |||  .
Db   6117 GTGCGCAATGGTTCCACCTGGCTGACGGCCTATGTGCTGCGCTCGTTTAGCAAGATCAAG 6176

Qy    522 SerIleValThrLysThrTyrGln------------GlyTrpGluGlyValThrAsnVal 537
          |||:::           |||            |||:::|||   :::
Db   6177 GATATAATTGACTTGGATGAGCAGATCCTGGCCAAGGGCTACGAATTCCTGTTAACTCGA 6236

Qy    538 AsnThrProPheGlyGlnPheAlaHisAlaGly---------LeuLeuLysAsnGluGlu 554
                  |||   |||        |||
Db   6237 CAGGCAGAAAATGGTAGCTTCACGGAACACGGAGAGTATTTCTATAGCTCCCAGCGATCT 6296

Qy    555 IleLeuCysLeu---AlaAspAspLeuAlaThrArgLeuThrGlyValTyrPro------ 571
          :::|||   |||   |||:::   |||         |||            |||
Db   6297 CTATTGACATTGACCGCAAATTCCTTGCTCGCCCTGCTGGAGGAGGAGAAACCGAACCAG 6356

Qy    572 ---AlaThrAspAsnPheAlaAlaAlaValSerAlaPheAlaAlaAsnMetLeuSerSer 590
             |||   |||         |||   :::||||||      |||      :::
Db   6357 GCTGCCATTGACAAGGCGGTTGCCTATCTGAGCGCCAATACTGCCGAGTCCATTGAGCTC 6416

Qy    591 ValLeuLysSerGluAlaThrSerSerIleIleLysSerValGlyGluThrAlaValGly 610
          :::   ||||||   |||      ::::::   |||:::           |||
Db   6417 CTGCCCAAATCCATTGCCATTTATGCCCTGCAAAAGGCCAAGGCGCCGGAAGCTGCCAAG 6476

Qy    611 ------AlaAlaGlnSerGlyLeuAlaLys 618
                |||:::        |||||||||
Db   6477 CAGGTGGCCAGTCTTAAGTCCCTGGCCAAA 6506

RESULT 11
AAX13087
ID   AAX13087 standard; DNA; 11739 BP.
XX
AC   AAX13087;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:150.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 862-868; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 11739 BP; 3917 A; 2161 C; 2537 G; 3118 T; 6 other;

Alignment Scores:
Pred. No.:                4.95          Length:          11739
Score:                    129.00        Matches:         160
Percent Similarity:       31.86%        Conservative:    107
Best Local Similarity:    19.09%        Mismatches:      255
Query Match:              3.82%         Indels:          317
DB:                       20            Gaps:            43

US-09-677-653A-50 (1-647) x AAX13087 (1-11739)

Qy      8 SerGlnArgProHisAsnArgArgGlyThrArgAsnValArgValSerAlaAsnThrVal 27
          :::   |||      |||       :::   |||:::      |||   |||   |||
Db   9107 ACGAATCGTAGTTCTAATAGCGGAGTAAGCACCAATATTATGTTTAGTAACAATCTTGTA 9166

Qy     28 ThrValAsnGlyArgArgAsnGlnArg---------------------------------- 36
                 |||:::   |||:::
Db   9167 GGAACAACAGGACAGTTCAATAAAGGGAATTCTGGTGCTACTATTAATTATTATATTTAT 9226

Qy     37 ----------------------------ArgArgThrGlyArgGlnValSerProProAsp 47
                                       |||:::::::::::::||||||
Db   9227 TACAAAAAGCTTTATGAGAATTTTGTTAATAGTAACGGACAAAAAATTACACCACCATCA 9286

Qy     48 AsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeuAspAlaAsnThrValThrPhe 67
             ||||||         :::   :::         :::::::::         ||||||
Db   9287 GGATTCACACAAGGAAAGCGAACGGTT---------ATTAATAGTGAAGCGTACACTTTT 9337

Qy     68 ProAlaAsnIleSerSerMetProGlu------------------------------Phe 77
                 :::      ::::::|||:::                             |||
Db   9338 AAACAAAGT---GGCACCTTGCCAGATACCTATCAAGCAGACGGCAAAACCTATAAGTTC 9394

Qy     78 ArgAsnTrpAlaLysGlyLysIle----------------------------------- 85
             :::   |||   |||||||||
Db   9395 AAAGGTTGGTACAAAGGCAAAACCAAGCCAAACACCTTGACCACCACCAAAGCACCAAGC 9454

Qy     86 ---------------------------AspLeu---------------------------- 87
                                       ||||||
Db   9455 TATGCGGTGACTTATAACAATGATGATGATTTGACGGTGGTGTATGAGGAGATTGAGTTT 9514

Qy     88 ---AspSerAspSerIleGlyTrpTyrPheLysTyrLeuAspProAlaGlyAlaThrGlu 106
             |||      ::::::   :::   |||   ::::::|||   :::|||
Db   9515 TTTGACTTCCCAGCGCTGACCTATCAATTCGGGTTCGTGGACGAGTCTGGCAAACGAGTA 9574

Qy    107 SerAlaArgAlaValGly---GluTyrSerLysIleProAspGlyLeuValLysPheSer 125
             |||      :::|||      |||   |||            |||:::      :::
Db   9575 GATGCTTCAACAATTGGGTTGACTTATGACAAATGGCGTGGGGAGTTGTTAACAAATAAT 9634

Qy    126 ValAsp-------------------------------------------------AlaGlu 129
          |||:::
Db   9635 GTCAATGACTGGAAAACAGTCAGTCTTGAAAAAGGTCAAGTTGCACCAACCAAAAACAAT 9694

Qy    130 IleArgGluIleTyr------AsnGluGluCysProValValThrAspValSerValPro 147
          :::::||||||          :::          ::::::::|||   |||
Db   9695 TTGAAGGAAATTGCTTATCCTGCGCAATCTCTTGAAATACTGAGTGATCGAAGTACA--- 9751

Qy    148 LeuAspGlyArgGlnTrpSerLeuSerIlePheSerPheProMet-------------- 162
                  |||:::|||   :::      :::|||   :::
Db   9752 -----------CAATATAGTGCGGCTAATTTAACCTTTACACTGCCCAAATACTATGGG 9799

Qy    163 ----------------------------PheArgThrAlaTyrValAlaValAlaAsn 172
                                         |||         |||:::         |||
```

```
Db    9800 AAAATCAGTGTTTATAACAAGAGTGGTACTTTTGATGCAGGCTACCTG---TTTCCAAAT 9856

Qy     173 ValGluAsnLysGluMetSerLeuAspValValAsnAspLeuIleGluTrpLeuAsnAsn 192
Db    9857 ATCCAAAATAATAATAGTTCAGAACCATCTGTTGAAGGCTTACACAAA---------AAT 9907

Qy     193 LeuAlaAspTrp---------------------------------------------- 196
Db    9908 CTTTCTAGCTGGTTTGAACTAAAGAAAAGTGGGAACCAATCGTTTGTTTTTAACAAACCG 9967

Qy     197 ----------------------------------------ArgTyrValValAspSer 202
Db    9968 ATTCCAGCAGCACCAATCGATGTTCAGGTGCCTGCTTACTTGCGAGAGATTGTCTATAAT 10027

Qy     203 GluGlnTrpIleAsnPheThrAsnAspThrThrTyrTyrValArgIleArgValLeuArg 222
Db   10028 CCCGATCTGATA------TCAAATACAGCCGTCTATTATGTA---------ATAGATAAA 10072

Qy     223 ProThrTyrAspValProAspProThrGluGlyLeuValArgThrValSerAspTyrArg 242
Db   10073 CCTGTCTAC-----------------------------------------TACTAT 10087

Qy     243 LeuThrTyrLysAlaIleThr-----------------------CysGluAlaAsnMet 254
Db   10088 TTAACCAACCGCAAAGTCACCGAAAACTTCGTAGACACAAGCGGTGCCAAAATTACACCA 10147

Qy     255 ProThrLeuValAspGlnGlyPheTrpIleGlyGlyGlnTyrAlaLeuThrPro-ThrSe 274
Db   10148 CCAACA-----------GGTTTCACCCAAGGAAAACAAACAGCGATTACCAGTGACCCA 10195

Qy     274 rLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHisThrLeuThrPheAlaArgPr 294
Db   10196 TACACCTTCA---AACAAGCAGGTACCTTACCTGATACCTACACAACAGGCGGTAAGACC 10252

Qy     294 oSerSerAlaAlaAlaLeuAlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaPr 314
Db   10253 TACAAGTTCAAAGGCT--------------------------------GGTACA----- 10274

Qy     314 oAlaGlyThrProAlaTrpGlu--------------GlnAlaSerSerGlyGlyTyrLe 329
Db   10275 -AAGGGAAAACCAAAACC-GAGCCTTTAACCACGACGAAAGCACCAAGTTATGGG---GT 10329

Qy     329 uThrTrpArgHisAsnGly-----------------------------ThrThrPhePr 339
Db   10330 AACCTATAATGATAATGATGATTTGAATGTGGTGTATGAAGAAGTAACAGCTATTTATCC 10389

Qy     339 oAlaGlySerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspProAs 359
Db   10390 AGGGTTCTATGCGCAATTTGTTGATGAACAAGGCAAAGCTTTTACTAATTCGCTAACTCT 10449

Qy     359 nAspGlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArg-------- 376
Db   10450 TTCTGAAAATTATACTGAGTTTCTCAGAAAGACAGGCACTACTGTTTTTGAAACGAAGGG 10509

Qy     377 -------------GlnValAlaValAspGluValValValThrAsnAsnProAlaGlyGl 392
Db   10510 CGCCTTTTACCCAATGGTAGGTAAGCGCGATGCAACAGTGACAAATAAGTATAAG----- 10564

Qy     392 yGlySerAlaProThrPheThrValArgValPro---------------------ProSe 405
Db   10565 ----ATTGAAACAACGAGTAGTGTTGCAGTACCAAATAATTATTGGACAGCCATGCCTGC 10620

Qy     405 rAsn------------------AlaTyrThrAsnThrValPheArgAsnThrLeu----- 417
Db   10621 TGACATTGCTAATAGAGGGTACGCATTAAATAACTTGGATATTACCAATCAAATGAAATA 10680

Qy     418 ----------LeuGluThrArgProSerSerArgArgLeu------------------- 427
Db   10681 TGTTGACAAGATAGAAACTACTGACACGCACACTACAATGTACGATTATACTGGAGTTTC 10740

Qy     428 -GluLeuProMetProProAlaAspPheGlyGlnThrValAlaAsnAsnProLysIleGl 447
Db   10741 GGACATTGGGGGCAAACCAAACAATGCTTCTCAGACAGAATTAAAT------CCTATTGA 10794

Qy     447 uGlnSerLeuLeuLysGluThrLeuGlyCysTyrLeuValHisSerLysMetArgAsnPr 467
Db   10795 TAAAACATTTGTTAAGCAAACAGATAACACTTTTGCGATAAAATCAATTTGGGGAAATGG 10854

Qy     467 oValPheGln-----------------------LeuThrProAlaSerSerPheGlyAl 479
Db   10855 AACGTTTTCGCATATAGTTCCCAAAATGTTATTGTTAGCAGAAAGTGGCAGTTTTTCAAG 10914

Qy     479 aVal---SerPheAsn---AsnProGlyTyrGluArgThrArgAspLeuProAspTyrTh 497
Db   10915 ATTATTCGGGTTTGATGGTACTCCAGACTACAAGCAAACC---------GTAAACTACAA 10965

Qy     497 rGlyIleArgAspSerPheAspGlnAsnMetSerThrAla---------------------- 510
Db   10966 AGTCACCCGTAAACAAGTCACCGAAAACTTTGTTGATGCCAACGGTGCCAAAATCACCGC 11025

Qy     510 ------------------------------------------------------------ 510
Db   11026 ACCAACAGGCTTCACCCAAGGCAACAAAATACCCATGACTAGCAACACCTTCAAGTACAC 11085

Qy     511 -ValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLysThrTyrGlnGl 530
Db   11086 AGCCGCAAAAGCCTTGCCAGCAACTTATACCGCAGGTGGCAAGACCTATACTTTCCAAGG 11145

Qy     530 yTrp---GluGlyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLe 549
Db   11146 CTGGTACAAAGGCAAAACGAAGCCTAATACT---------------------------- 11176

Qy     549 uLeuLysAsnGluGluIleLeuCysLeuAlaAspAspLeuAlaThrArgLeuThrGlyVa 569
Db   11177 ------------------------------------TTAACAACCAGCACTACGCCAAC 11199

Qy     569 lTyrProAlaThrAspAsnPheAlaAlaAlaValSerAlaPheAlaAlaAsn---MetLe 588
Db   11200 ATATAACACAACT-------------------------TTTGATGGCAATGACGACAT 11232

Qy     588 uSerSerValLeuLysSerGlu---------AlaThrSerSerIleIleLysSerValGl 605
Db   11233 GACGGCGATTTATAAGGAAGAAGTGCCCAAAGCAAGTGTCGCTTTAACGCGCACTACCGC 11292

Qy     605 yGluThrAla-ValGlyAlaAlaGlnSer---GlyLeuAlaLysLeuPro 620
Db   11293 TGAAACCGTGACTAGCGGCGGCAATGTCACCTGGCGCGCCACAATTACCA 11342
```

RESULT 12
ABL19215
ID ABL19215 standard; DNA; 6592 BP.
XX
AC ABL19215;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 9118.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX

```
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 9118; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6592 BP; 1705 A; 1969 C; 1673 G; 1245 T; 0 other;

Alignment Scores:
Pred. No.:                3.29          Length:          6592
Score:                    127.00        Matches:         77
Percent Similarity:       32.80%        Conservative:    45
Best Local Similarity:    20.70%        Mismatches:      116
Query Match:              3.76%         Indels:          134
DB:                       23            Gaps:            18

US-09-677-653A-50 (1-647) x ABL19215 (1-6592)

Qy     231 ThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAlaIleThrCys 250
           ||||||         :::   |||      |||:::::::::::::::|||
Db    4294 ACTGAGACAAAGACCCAGAAGGTTCCCCACTACCAGATGAGTTTCGAGGCCACCGAACTG 4353

Qy     251 GluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIle------------------ 264
           |||   |||   |||                 ||||||:::
Db    4354 GAGAAGAACAAGCCC------------TACGAGTTCTGGGTGACAGCTAGCACCACCATT 4401

Qy     265 ---GlyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAla 283
               ||||||   :::   :::      :::::::|||         |||
Db    4402 GGCGAAGGCCAGCAGTCTAAGAGCATAGTGGCCATGCCCAGCGACCAGGTGCCCGCCAAG 4461

Qy     284 TyrAla------------LeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAla 299
             |||               |||   ||||||         :::
Db    4462 ATCGCCTCCTTCGACGACACCTTCACTGCCACCTTCAAGGAGGACGCCAAGATGCCCTGC 4521

Qy     300 LeuAlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAla 319
           ||||||           |||   ||||||
Db    4522 CTGGCC---------GTTGGAGCCCCCCAA---------------------------- 4542

Qy     320 TrpGluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhePro 339
                                  :::|||||:::      |||      |||
Db    4543 --------------------CCCGAGATCACATGGAAGATCAAGGGCGTCGAATTCAGT 4581

Qy     340 AlaGlySerValSerTyrValLeuProGluGlyPheAlaLeu------------GluArg 355
           |||               |||||||||:::|||      |||              |||
Db    4582 GCCAACGATCGCATGCGCGTCCTCCCCGACGGATCGCTGCTGATCAAGTCGGTAAATCGC 4641

Qy     356 TyrAspProAsnAspGlySer---TrpThrAspPheAlaSerAlaGlyAspThrValThr 374
             |||      |||   |||         :::   :::   |||   |||:::::::|||
Db    4642 CAGGATGCCGGAGACTACTCCTGCCACGCCGAGAACTCGATTGCTAAGGACTCCATCACG 4701

Qy     375 PheArgGlnValAlaVal------------------------------------------ 380
             :::   :::   :::
Db    4702 CACAAACTGATTGTTCTGGCACCACCACAATCGCCCCACGTCACCCTCTCGGCAACCACC 4761

Qy     381 ---AspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThr------ 397
               |||   :::   |||         |||      |||   :::|||||||
Db    4762 ACTGATGCCCTGACTGTAAAGTTGAAGCCCCATGAGGGAGACACTGCTCCTCTGCATGGA 4821

Qy     398 PheThrValArgValProProSer------------------------------------ 405
           :::|||:::         |||
Db    4822 TACACCCTGCACTACAAGCCAGAATTCGGAGAATGGGAAACATCGGAAGTGTCTGTCGAC 4881

Qy     406 ----------------------------------------------AsnAlaTyrThrAsn 410
                                                           |||
Db    4882 TCACAGAAGCACAACATCGAAGGTCTCTTGTGCGGCTCCCGCTATCAGGTCTATGCC--- 4938

Qy     411 ThrValPheArgAsn--------------------ThrLeuLeuGluThrArgProSer 423
           |||   |||   |||                    :::|||   ||||||
Db    4939 ACAGGATTCAATAACATTGGAGCTGGCGAAGCTTCTGACATTTTGAACACCCGGACCAAG 4998

Qy     424 SerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrValAlaAsnAsn 443
               ::::::   :::||||||      |||      |||   :::      :::|||:::
Db    4999 GGACAGAAGCCCAAGCTGCCCGAGAAACCTCGC---TTCATTGAAGTGTCGTCCAACTCG 5055

Qy     444 ProLysIleGluGlnSerLeuLeuLysGluThrLeuGlyCys------------TyrLeu 459
                  :::                |||:::      ||||||           ::::::
Db    5056 GTGTCCCTGCACTTCAAGGCCTGGAAAGATGGA---GGATGCCCCATGTCGCACTTTGTG 5112

Qy     460 ValHisSerLysMetArgAsnPro----------------------------------- 467
           |||   ||||||   |||:::
Db    5113 GTGGAGAGCAAGAAGCGCGATCAAATTGAGTGGAACCAAATTTCGAACAACGTGAAGCCC 5172

Qy     468 ---------------ValPheGlnLeuThrProAlaSerSerPheGly------AlaVal 480
                           |||      |||   ||||||:::   :::
Db    5173 GATAACAACTACGTGGTTTTGGACCTGGAACCCGCCACCTGGTACAACCTCCGCATCACT 5232

Qy     481 SerPheAsnAsnProGlyTyr---------------------------------------- 487
           :::   |||:::   |||:::
Db    5233 GCCCACAACTCGGCTGGCTTCACTGTGGCCGAATACGACTTTGCCACCTTAACCGTTACC 5292

Qy     488 ------------GluArgThrArgAspLeuProAsp 495
                              :::|||||||||||||:::
Db    5293 GGAGGCACTATCGCACCCTCGCGAGATTTACCCGAG 5328

RESULT 13
AAS31116
ID   AAS31116 standard; cDNA; 8277 BP.
XX
AC   AAS31116;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human diagnostic and therapeutic polynucleotide (DITHP) #131.
XX
KW   Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW   cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW   acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW   respiratory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200162927-A2.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-US06059.
XX
PR   24-FEB-2000; 2000US-0184693.
PR   24-FEB-2000; 2000US-0184697.
PR   24-FEB-2000; 2000US-0184698.
PR   24-FEB-2000; 2000US-0184768.
PR   24-FEB-2000; 2000US-0184769.
PR   24-FEB-2000; 2000US-0184770.
PR   24-FEB-2000; 2000US-0184771.
PR   24-FEB-2000; 2000US-0184772.
PR   24-FEB-2000; 2000US-0184773.
PR   24-FEB-2000; 2000US-0184774.
PR   24-FEB-2000; 2000US-0184776.
PR   24-FEB-2000; 2000US-0184777.
PR   24-FEB-2000; 2000US-0184797.
PR   24-FEB-2000; 2000US-0184813.
PR   24-FEB-2000; 2000US-0184837.
```

PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204525.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'Sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
DR WPI; 2001-502867/55.
DR P-PSDB; AAU19545.
XX
PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
PT -
XX
PS Claim 1; Page 356-358; 522pp; English.
XX
CC The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and
CC therapeutic (DITHP) polynucleotides of the invention.
XX
SQ Sequence 8277 BP; 2316 A; 1962 C; 1890 G; 2106 T; 3 other;

Alignment Scores:

Pred. No.:	4.96	Length:	8277
Score:	126.50	Matches:	133
Percent Similarity:	34.87%	Conservative:	71
Best Local Similarity:	22.74%	Mismatches:	186
Query Match:	3.75%	Indels:	195
DB:	22	Gaps:	33

US-09-677-653A-50 (1-647) x AAS31116 (1-8277)

```
Qy      12 HisAsnArgArgGlyThrArgAsnValArgValSerAlaAsn-ThrValThrValAsnGl 31
           ||||||   :::         ::::::   |||   |||||| |||
Db     997 CACAATGCCAAGTGCCAGCTGTCCCTCGAGGTGTTCGCCAACGACA---------- 1042

Qy      31 yArgArgAsnGln-----------ArgArgArgThrGlyArgGlnValSerProProAs 47
            ||||||:::              |||   ||||||              ||||||
Db    1043 -AGGAGATCTGCATGATCAAGGTAGAGATCCAGGACATCAACGACAACGCGCCCTCCTTC 1101

Qy      47 pAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeu---------- 60
           |||   :::              ::::::|||
Db    1102 TCCTCGGACCAGATCGAAATGGACATCTCGGAGAACGCTGCTCCGGGCACCCGCTTCCCC 1161

Qy      61 ----AspAlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAs 79
               |||:::           ||||||   :::   |||   |||
Db    1162 CTCACCAGCGCACATGACCCCGACGCCGGCGAGAATGGGCTCCGCACCT---------AC 1212

Qy      79 nTrpAlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTrpTyrPheLysTyrLe 99
           ::::::      |||         :::   |||:::   |||         :::
Db    1213 CTGCTCACGCGCGACGATC---------ACGGCCTCTTTGGACTGGACGTTAAGTCACGC 1263

Qy      99 uAspProAlaGlyAlaThrGluSerAlaArgAlaValGlyGluTyrSerLysIleProAs 119
           |||                                                :::   |||::
Db    1264 GGCGACGGCA----------------------------------CCAGAGTTCCCAGA 1287

Qy     119 pGlyLeuValLysPheSerValAspAlaGluIleArgGluIleTyrAsnGluGluCysPr 139
           :   :::::::::   ::::::|||         ||||||      |||
Db    1288 ACTGGTCATCCAGAAGGCTCTGGAC---------CGCGAGCAACAGAATCACCATACGCT 1338

Qy     139 oValValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSe 159
           |||:::|||              |||||||||   :::
Db    1339 CGTGCTGACT-----------GCCCTGGACGGTGGCGAG-------------------- 1366

Qy     159 rPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMetSe 179
           |||         |||      :::   |||      :::::::::   :::   ||
Db    1367 -CCTCCACGTTCCGCCACCGTACAGATCAACGTGAAGGTGATTGACTCCAACGACAACAG 1425

Qy     179 r------LeuAspValValAsnAspLeuIleGluTrpLeuAsnAsnLeuAlaAspTrpAr 197
           |      :::      :::   |||:::|||        |||
Db    1426 CCCGGTCTTCGAGGCGCCATCCTACTTGGTGGAACTGCCCGAGAACACTCCGCTGGGTAC 1485

Qy     197 gTyrValValAspSerGluGlnTrpIleAsnPheThrAsn---------- 210
           |||:::|||           :::|||   |||:::
Db    1486 AGTGGTCATCGAT-----------CTGAACGCCACCGACGCCGATGAAGGTCCCAATGG 1533

Qy     211 -AspThrThrTyr-----------TyrValArgIleArgValLeuArgProThrTyrAs 226
            :::      |||           ||||||      ||||||   |||      :::
Db    1534 TGAAGTGCTCTACTCTTTCAGCAGCTACGTGCCTGACCGCGTG---CGGGAGCTCTTCTC 1590

Qy     226 pValProAspProThrGluGlyLeuValArgThrValSer------AspTyr-------- 241
           :::   ||||||      |||||:::|||              ||||||
Db    1591 CATC---GACCCCAAGACCGGCCTAATCCGTGTGAAGGGCAATCTGGACTATGAGGAAAA 1647

Qy     241 ---------------------------------------------------------- 241

Db    1648 CGGGATGCTGGAGATTGACGTGCAGGCCCGAGACTTGGGGCCTAACCTTATCCCAGCCCA 1707

Qy     242 ----ArgLeuThrTyrLysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGl 260
               ::::::|||   |||   |||           |||   |||::::::
Db    1708 CTGCAAAGTCACGGTCAAGCTCATCGACCGCAACGACAATGCGCCGTCCATC-------- 1759

Qy     260 nGlyPheTrpIleGlyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspVa 280
           ||||||                        |||:::   |||      ::
Db    1760 -GGTTTC----------------------------GTCTCCGTGCGCCAGGGGGCGCT 1788

Qy     280 lSerGluAlaTyrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLe 300
           :|||||||||                        |||   |||   :::      |||||
Db    1789 GAGCGAGGCC----------------------GCCCCTCCCGGCACCGTCATCGCCCT 1824

Qy     300 uAlaPheValTrp---------AlaGly--------------------------LeuPr 308
           |      |||           :::|||                           |||
Db    1825 GGTGCGGGTCACTGACCGGGACTCTGGCAAGAACGGACAGCTGCAGTGTCGGGTCCTAGG 1884
```

```
Qy     308 oGlnGlyGlyThrAlaProAlaGlyThrProAlaTrpGluGlnAlaSerSerGlyGlyTy 328
           |||||||||       |||
Db    1885 CGGAGGAGGGACGGGCGGCGGCGGG------------------------------------ 1909

Qy     328 rLeuThrTrpArgHisAsnGlyThrThrProAlaGlySerValSerTyrValLeuPr 348
                       |||         |||   |||||||||   :::   |||
Db    1910 -------------------GGCCTGGGCGGGCCCGGGGGTTCCGTCCCCTTCAAGCTTGA 1950

Qy     348 oGluGlyPheAlaLeuGluArgTyrAspProAsnAspGlySerTrpThrAspPheAlaSe 368
            |||   :::                    |||     :::|||
Db    1951 GGAGAACTAC----------------------GACAACTTCTACACG----------- 1975

Qy     368 rAlaGlyAspThrValThrPheArgGlnVal---------AlaValAspGluValValVa 385
                   ||||||   |||   :::             ||||||      ||
Db    1976 ----------GTGGTGACTGACCGCCCGATGGACCGCGAGACACAAGACGAGTACAACGT 2025

Qy     385 lThrAsnAsnProAlaGlyGlyGlySerAlaPro---------------ThrPheThrVa 400
           ||||               |||||||||   |||               :::|||   ::
Db    2026 GACCATCGTGGCGCGGGACGGGGGCTCTCCTCCCCTCAACTCCACCAAGTCGTTCGCGAT 2085

Qy     400 lArgVal-----------------ProPro-SerAsnAlaTyrThrAsnThrValPheA 414
           :::::::                ||||||   :::         :::|||   ||||
Db    2086 CAAGATTCTAGACGAGAACGACAACCCGCCTCGGTTCACCAAAGGGCTCTACGTGCTTCA 2145

Qy     414 rgAsnThrLeuLeuGluThrArgProSerSerArgArgLeuGluLeuProMetProProA 434
           ||  |||         :::|||  |||:::     |||     |||      |||
Db    2146 GGTGCACGAGAACAACATCCCGGGAGAGTACCTGGGCTCTGTGCTCGCCCAGGATCCCG- 2204

Qy     434 laAspPheGlyGlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluT 454
                  |||   |||   |||      ::      |||:::|||
Db    2205 -------------ACCTGGGCCAAGAACGGCACCGTATCCTACTCTATCCTGCCCTCGC 2250

Qy     454 hrLeuGlyCysTyrLeuValHisSerLysMetArg---AsnProValPheGlnLeuThrP 473
             :::|||        :::::::::  :::     ||||||          |||
Db    2251 ACATCGGCGACGTGTCTATCTACACCTATGTGTCTGTGAATCCC-----------ACGA 2298

Qy     473 roAlaSerSerPheGlyAlaValSerPheAsnAsnProGlyTyrGluArgThrArgAspL 493
           :::   :::        |||||||||        :::|||:::|||:::
Db    2299 ACGGGGCCATCTACGCCCTGCGCTCCTTTAAC---------TTCGAGCAGACCAAGGCTT 2349

Qy     493 euProAspTyrThrGlyIleArgAspSerPheAspGlnAsnMetSerThrAlaValAlaH 513
                       :::|||                  |||   |||   ||||
Db    2350 TTGAGTTCAAGGTGCTTGCTAAGGAC----------------TCGGGGGCGCCCGCGC 2391

Qy     513 isPheArgSer 516
           ||      |||
Db    2392 ACTTGGAGAGC 2402

RESULT 14
ABQ90255
ID   ABQ90255 standard; DNA; 1548 BP.
XX
AC   ABQ90255;
XX
DT   01-OCT-2002  (first entry)
XX
DE   M. capsulatus gene #240 for DNA array.
XX
KW   Micro array; gene; ds; differential expression; gene expression.
XX
OS   Methylococcus capsulatus.
XX
PN   WO200255655-A2.
XX
PD   18-JUL-2002.
XX
PF   14-JAN-2002; 2002WO-NO00019.
XX
PR   12-JAN-2001; 2001NO-0000235.
PR   12-JAN-2001; 2001NO-0000239.
XX
PA   (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA   (TIGR-) TIGR.
XX
PI   Birkeland NK,  Eidhammer I,  Jonassen I,  Jensen HB,  Lien T;
PI   Lillehaug JR,  Lossius I,  Eisen JA,  Fraser CM,  Durkin AS;
PI   Salzberg SL;
XX
DR   WPI; 2002-557818/59.
XX
PT   Novel DNA array useful for determining differential expression of
PT   Methylococcus capsulatus genes, comprises polynucleotides or
PT   oligonucleotides representative for a selective number of Methylococcus
PT   capsulatus genes  -
XX
PS   Claim 19; Page 166; 678pp; English.
XX
CC   The invention relates to a novel DNA array giving a representation of a
CC   number of Methylococcus capsulatus genes. The method of the invention is
CC   useful for determination of the differential expression of the genes of
CC   M. capsulatus, and for studying gene expression on a genomic scale and in
CC   gene expression assays of M. capsulatus genes. The sequences shown in
CC   ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
CC   invention.
XX
SQ   Sequence 1548 BP; 231 A; 531 C; 457 G; 329 T; 0 other;

Alignment Scores:
Pred. No.:              0.543          Length:              1548
Score:                  126.00         Matches:             127
Percent Similarity:     32.15%         Conservative:        64
Best Local Similarity:  21.38%         Mismatches:          187
Query Match:            3.73%          Indels:              218
DB:                     24             Gaps:                31

US-09-677-653A-50 (1-647) x ABQ90255 (1-1548)

Qy     141 ValThrAspVal---SerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSer 159
           :::|||:::::   :::|||||||||      |||   |||         |||
Db       1 ATGACCGAGATCGAGAACGTCCCTCTC------CGCGCCTGG---------ATCGGCTTT 45

Qy     160 PheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMetSer 179
           |||   |||           :::::::|||:::   :::::::::             |||
Db      46 TTCGGGATGGTGTTGGGAGTCTTCATGGCGATCTTGGACATCCAGATCGTCGCCAGCTCG 105

Qy     180 LeuAspValVal--------------------AsnAspLeuIleGluTrpLeuAsnAsn 192
           |||:::   :::                    |||   |||   |||:::
Db     106 CTGGAGCAGATCCAGGCCGGCCTCGCCGCCACCCGCGACGAGATTACCTGGGTGCAGACC 165

Qy     193 ------LeuAlaAspTrpArgTyrValValAspSerGluGlnTrpIleAsnPheThrAsn 210
                 |||               |||:::           |||::::||
Db     166 GCTTACCTGATCGCCGAAGTGGTCGTCATCCCGCTGTCCGGGTGGCTGAACCGCGCCCTC 225

Qy     211 AspThrThrTyrTyrValArgIleArgValLeuArgProThrTyrAspValProAspPro 230
               |||   |||
Db     226 TCGACCCGCTACCTG---------------------------------------------- 240

Qy     231 ThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAlaIleThrCys 250
                     |||   :::         :::   ||||||
Db     241 ---------TTCGTGTCGTCCTGCGGCGGTTTCACCTTGACCAGCCTGTGCTGCGCGCTG 291

Qy     251 GluAlaAsnMetProThrLeuValAsp-----------GlnGlyPheTrpIleGlyGly 266
                ||::::|||::::::::                |||||||||:::   ||||||
Db     292 GCGTGGAACCTCCCCTCGATGATCGCGTTCCGGGCCCTGCAGGGCTTCTTC---GGCGGG 348

Qy     267 GlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeu 286
                ||:::   ||||||                                      :::
Db     349 ------GCCATGATCCCGACCGTATTCGCG----------------------GTGATT 378

Qy     287 HisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeuAlaPheValTrpAlaGly 306
           :::||||||           |||           |||         :::      |||
Db     379 TACACCCTGTTC---------CCGCCCCGCCAGCAGGCGACCATGGTGATCGTGGTCGGC 429

Qy     307 LeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrpGluGlnAlaSerSerGly 326
           |||               ||||||         |||                     |||
Db     430 CTGGTGGTG---ACGGTGGCCCCCACCGCCGGCCCCGTG--------------CTCGGC 471
```

```
Qy    327 GlyTyrLeuThr------------TrpArgHis-------------------------- 333
          ||||||||||||           |||:::
Db    472 GGCTATCTGACCGAAGTCCTGTCCTGGAAGGCCCTCTTCCTGGTCAATCTGCTGCCGGGC 531

Qy    333 ------------------------------------------------------------ 333

Db    532 GTGCTGGCCTGCCTTTCGACCTGGTTCCTCGTGGAGGTAGACCAGCCGGAATGGACGCTG 591

Qy    334 ---AsnGlyThrThrPheProAla-----------------------GlySerValSer 344
                      ||||||                            ||||||:::
Db    592 CTCAAACGCGTCGATTTTCCCGGCATCCTCTACATCGGCGTTTTTCTGGGCAGCCTCCAA 651

Qy    345 TyrValLeuProGluGlyPheAlaLeuGluArgTyrAspProAsnAspGlySerTrpThr 364
          :::||||||   ||||||        |||   :::|||      :::      :::
Db    652 TTCGTCCTGGAAGAAGGCATCCGCGAGGAATGGTTCGATAGCCGAGAGATCGTGTTCTTC 711

Qy    365 AspPheAlaSer------AlaGlyAspThrValThrPheArgGlnValAlaValAspGlu 382
                :::      ||||||      :::   :::|||:::::   :::::::|||
Db    712 AGCGCCGTCGCGGTAATGGCCGGCGGCGCCATGTTCTACCGCGAGCTGCGCATCGAGGAA 771

Qy    383 ValValVal--------------ThrAsnAsnProAlaGlyGlyGlySerAlaProThr 397
             :::|||                  |||                 |||       :::
Db    772 CCCATCGTCGACCTGTGGGCGTTCCGTAACACCAATTTCGCGGTCGGCTGCCTGTTCAGC 831

Qy    398 PheThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeu 417
          |||   :::  :::                 ||||||
Db    832 TTCATCATCGGCATC---------GGGCTCTATACCCTCATCGCGTTGACGCCCGTCTAC 882

Qy    418 LeuGluThrArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGly 437
          |||   |||                   |||:::
Db    883 CTTTCCACGGTGAAAGGCCTCAACAGCCTGCAG-------------------------- 915

Qy    438 GlnThrValAlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeuGlyCys 457
                                                              :::|||
Db    916 ------------------------------------------------ATCGGGCAG 924

Qy    458 TyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPhe 477
          ||||||:::   :::   |||           ||||||        ||||||:::|||
Db    925 TACCTGATGGTCACCGGCATG-----------TTCCAGTTC------GCCTCGGCTTTC 966

Qy    478 GlyAlaValSerPheAsnAsnProGlyTyrGluArgThrArgAspLeuProAspTyrThr 497
             |||               |||      :::|||
Db    967 GTGGCG-----------GGACCCTTGGCGAAGCGG----------------------- 990

Qy    498 GlyIleArgAspSerPheAspGlnAsnMetSerThrAlaValAlaHisPhe--------- 514
                         |||      :::      |||:::      |||
Db    991 --------------ATGGACTCGCGTCTGATGTTGGCGCTGGGCCTCTTCCTCTTCGGA 1035

Qy    515 ------------------ArgSerLeuSerHisSerCysSerIleValThrLysThrTyr 528
                           ||||||||||||
Db   1036 CTGGGCTCATGGCTGAACGGAAGCCTGAGCCACGAC----------------------- 1071

Qy    529 GlnGlyTrpGluGlyVal-----------------------------ThrAsnValAsn 538
          |||:::   |||:::                                  :::|||
Db   1072 TCCGGCTACTG-GGAATTCTTCGGGCCCCAGGCCCTGCGCGGCTTCGCCCTGATGTTCGC 1130

Qy    539 ThrProPheGlyGlnPheAlaHisAla-GlyLeuLeuLysAsnGluGluIleLeuCysLe 558
             |||      |||||||||   ||| |||   |||      ||||||:::
Db   1131 CTTCCTGCCCATCAATTCGCTGGCGCTGGGAACCCTGCCACCAGAGGAGGTC-------- 1182

Qy    558 uAlaAspAspLeuAlaThrArgLeuThrGlyValTyrProAlaThrAspAsnPheAlaAl 578
                              :::|||:::|||      |||   |||
Db   1183 ---------------AAGAATGCCAGCGGGCTGTACAACCTCACCCGCAATCTGGGCGG 1226

Qy    578 aAlaValSerAlaPheAlaAlaAsnMetLeuSer-----------------------Se 590
           |||:::           ||||||   |||                          ::
Db   1227 CGCCATCGGCATCGCCGTCGCCAATACCCTGATGATCCGCTGGGACAAAGCGCATTACGC 1286

Qy    590 rValLeuLysSerGluAlaThrSerSerIleIleLysSerValGlyGluThrAlaValGl 610
          :   |||:::        |||               |||   |||
Db   1287 CGCCCTGCGCGATTCCGTCACGCC-GGGTTC------TCCACAGGCCGACGCCCTGCTGA 1339

Qy    610 yAlaAlaGlnSerGlyLeuAlaLysLeuProGlyLeuLeuMetSerValProGlyLysIl 630
            ||||||   |||                                      ||||||:::
Db   1340 CGCGGCTGGAAGC-----------------------------------CCAGGCCGCCC 1363

Qy    630 eAlaAlaArgValArgAlaArgArgAlaArgArgArgAla 643
            |||   |||:::|||:::|||||||   |||   :::|||
Db   1364 TGCTCCGCGTCTCCGATCCCGCCGACGCCGCACTCAAGCA 1403
```

RESULT 15
AAX81820
ID AAX81820 standard; cDNA; 2748 BP.
XX
AC AAX81820;
XX
DT 02-SEP-1999 (first entry)
XX
DE Talaromyces emersonii glucoamylase enzyme gene locus.
XX
KW Glucoamylase enzyme; saccharification; starch hydrolysate;
KW dextrose syrup; syrup; ethanol; fuel; beverage; citric acid;
KW ascorbic acid; lysine; glutamic acid; ss.
XX
OS Talaromyces emersonii.
XX
PN WO9928448-A1.
XX
PD 10-JUN-1999.
XX
PF 26-NOV-1998; 98WO-DK00520.
XX
PR 10-JUL-1998; 98DK-0000925.
PR 26-NOV-1997; 97US-0979673.
PR 30-DEC-1997; 97DK-0001557.
PR 30-JUN-1998; 98US-0107657.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Lehmbeck J, Nielsen BR, Nielsen RI;
XX
DR WPI; 1999-404822/34.
DR P-PSDB; AAY23339.
XX
PT New glucoamylase obtained from Talaromyces emersonii
XX
PS Claim 13; Fig 13; 79pp; English.
XX
CC The present sequence represents a glucoamylase gene locus obtained from
CC Talaromyces emersonii. The glucoamylase enzymes have high thermal
CC stability so that a saccharification process may be carried out within
CC a shorter period of time or the process may be carried out using a lower
CC enzyme dosage. The glucoamylase enzymes can be used for saccharifying
CC starch hydrolysate for converting starch or partially hydrolysed starch
CC into a syrup containing dextrose. They can be used for producing
CC oligosaccharides, speciality syrups, ethanol for fuel, beverages or
CC organic compounds such as citric acid, ascorbic acid, lysine or glutamic
CC acid.
XX
SQ Sequence 2748 BP; 628 A; 815 C; 674 G; 631 T; 0 other;

Alignment Scores:

Pred. No.:	1.32	Length:	2748
Score:	125.50	Matches:	103
Percent Similarity:	31.94%	Conservative:	58
Best Local Similarity:	20.44%	Mismatches:	182
Query Match:	3.72%	Indels:	161
DB:	20	Gaps:	23

US-09-677-653A-50 (1-647) x AAX81820 (1-2748)

```
Qy     40 GlyArgGlnValSerProProAspAsn------------------------------- 48
          |||||||||::::::   ||||||
Db   1370 GGTAGGCAAATGAATATTCCCGACACAGCGTGGTACTAATTTGATTCAGACCTCTGGGAA 1429
```

```
Qy      49 --------------------PheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeuAsp 61
                               ||||||  |||  |||     |||     ::::::
Db    1430 GAAGTAGAAGGATCCTCATTCTTCACAACCGCCGTGCAACACCGCGCC---CTGGTCGAA 1486

Qy      62 AlaAsnThrValThrPheProAlaAsn-----------------IleSerSerMetPro 75
              |||   ::            |||                  :::|||      |||
Db    1487 GGCAATGCACTGGCAACAAGGCTGAACCACACGTGCTCCAACTGCGTCTCTCAGGCCCCT 1546

Qy      76 GluPhe--------------ArgAsnTrp---------------------------- 80
           :::               |||
Db    1547 CAGGTCCTGTGTTTCCTGCAGTCATACTGGACCGGATCGTATGTTCTGGCCAACTTTGGT 1606

Qy      81 AlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTrpTyrPheLysTyrLeuAsp 100
              |||:::        |||   :::|||||||                      |||
Db    1607 GGCAGCGGTCGTTCCGGCAAGGACGTGAATTCGATTCTGGGCAGCATCCACACCTTTGAT 1666

Qy     101 ProAlaGlyAlaThrGlu--------------------SerAlaArgAlaValGlyGlu 113
           |||||||||      :::                  |||||||||||||:::
Db    1667 CCCGCCGGAGGCTGTGACGACTCGACCTTCCAGCCGTGTTCGGCCCGTGCCTTGGCAAAT 1726

Qy     114 TyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArgGluIle 133
           :::      :::   |||                                  |||   |||
Db    1727 CACAAGGTGGTCACCGAC-------------------------TCGTTCCGGAGTATC 1759

Qy     134 TyrAsnGluGluCysProValValThrAspValSerValProLeuAspGlyArg------ 151
           |||              :::            :::|||   :::   ||||||
Db    1760 TATGCGATCAACTCAGGCATCGCAGAGGGATCTGCCGTGGCAGTC---GGCCGCTACCCT 1816

Qy     151 ------------------------------------------------------------ 151

Db    1817 GAGGATGTCTACCAGGGCGGGAACCCCTGGTACCTGGCCACAGCAGCGGCTGCAGAGCAG 1876

Qy     152 -----------------GlnTrp-----------------SerLeuSerIlePheSer 159
                             ||||||                 |||::::::      |||
Db    1877 CTTTACGACGCCATCTACCAGTGGAAGAAGATCGGCTCGATAAGTATCACGGACGTTAGT 1936

Qy     160 PheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMetSer 179
              |||   |||:::     |||   :::   |||   |||
Db    1937 CTGCCATTTTTCCAGGATATCTACCCTTCTGCCGCG---GTGGGCACCTATAACTCTGGC 1993

Qy     180 LeuAspValValAsnAspLeuIleGluTrpLeuAsnAsnLeuAlaAspTrp--------- 196
                    ||||||:::|||      :::          |||
Db    1994 TCCACGACTTTCAACGACATCATCTCGGCCGTCCAGACGTATGGTGATGGATATCTGAGT 2053

Qy     197 ---------ArgTyrValValAspSerGlu-----------GlnTrp-IleAsnPheTh 209
                   |||:::   :::|||||||:::            :::|||    |||
Db    2054 ATTGTCGTACGTTTTGCCTTAGATTCTCAGGTGTAAAGAAAAAAAATGGAACTAACTCAGT 2113

Qy     209 rAsnAspThrThrTyrTyrValArgIleArgValLeuArgProThrTyrAspValProAs 229
                  |||   :::|||           |||   :::|||
Db    2114 TCTAGGAGAAATATACTCCCTCAGA-----------CGGCTCTCTTACCGAACAATTCT 2161

Qy     229 pPro-ThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAla---I 248
            ||| |||:::|||         :::   |||            :::|||   :::   :
Db    2162 CCCGTACAGACGGCACTCCGCTTTCTGCCTCTGCCCTGACTTGGTCGTACGCTTCTCTCC 2221

Qy     248 leThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGlnT 268
           ::|||      |||         ::::::|||
Db    2222 TAACCGCTTCGGCCCGCAGACAGTCCGTCGTC--------------------------- 2253

Qy     268 yrAlaLeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyrAlaLeuHisT 288
                  |||   |||       :::          |||         |||:::   :
Db    2254 ----------CCTGCTTCCTGGGGCGAAAGCTCCGCAAGCAGCGTCCCTGCCGTCTGCT 2302

Qy     288 hrLeuThrPheAlaArgProSerSerAlaAlaAlaLeuAlaPheValTrpAlaGlyLeuP 308
           ::   |||   |||             :::   |||         ||||||
Db    2303 CTGCCACCTCTGCCACGGGCCCATACAGCACGGCTACCAACACCGTCTGGCCAAGCTCTG 2362

Qy     308 roGlnGly--------------GlyThrAlaProAlaGlyThrProAlaTrpGluGlnA 323
           |||                 :::|||||      ||||||
Db    2363 GCTCTGGCAGCTCAACAACCACCAGTAGCGCCCCATGCACCACTCCTACCTCTGTGGCTG 2422

Qy     323 laSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPheProAlaGlySerV 343
```

```
           :::           ::::::      :::      |||   :::   |||||||||:
Db    2423 TGACCTTCGACGAAATCGTCAGCACCAGTTACGGGGAGACAATCTACCTGGCCGGCTCGA 2482

Qy     343 alSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspProAsnAspGlySerT 363
           ::                                          |||      |||:::|
Db    2483 TC-----------------------------------------CCCGAGCTGGGCAACT 2500

Qy     363 rpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaValAspGluV 383
           ||            |||            :::      |||               :::
Db    2501 GG-----------TCCACGGCCAGCGCGATCCCCCTCCGCGCG-----------GATG 2536

Qy     383 alValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPheThrValArgValP 403
                     :::||||||                                |||||   :::|
Db    2537 CTTACACCAACAGCAACCCGCTCTGG-------------TACGTGACCGTCAATCTGC 2581

Qy     403 roProSerAsnAlaTyrThrAsnThrValPheArgAsn-----------ThrLeuLeuG 419
           |||||      ::::::           |||:::|||           |||::::::
Db    2582 CCCCTGGCACCAGCTTCGAGTACAAGTTCTTCAAGAACCAGACGGACGGGACCATCGTCT 2641

Qy     419 luThrArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnT 439
                     :::|||         :::         ||||||      |||||||
Db    2642 GGGAAGACGACCCGAACCGGTCGTACACGGTC---------CCAGCGTACTGTGGGCAGA 2692

Qy     439 hrValAla 441
           ||   |||
Db    2693 CTACCGCC 2700
```

Search completed: February 27, 2003, 05:13:00
Job time : 343 secs